UNIVERSITY OF ALBERTA

THE ASPARTIC PROTEINASES: STRUCTURE, INHIBITION, AND ZYMOGEN ACTIVATION

by

AMIR R. KHAN

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES AND RESEARCH

IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

DEPARTMENT OF BIOCHEMISTRY

EDMONTON, ALBERTA

SPRING, 1999

National Library of Canada
Acquisitions and Bibliographic Services
395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque nationale du Canada
Acquisitions et services bibliographiques
395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence

Our file Notre référence

The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

L'auteur conserve la propriété du droit d'auteur qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

0-612-39554-5

Canada

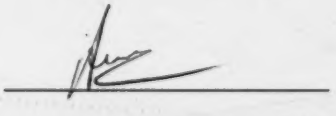
UNIVERSITY OF ALBERTA

LIBRARY RELEASE FORM

NAME OF AUTHOR:	Amir R. Khan
TITLE OF THESIS:	The Aspartic Proteinases: Structure, Inhibition, and Zymogen Activation
DEGREE:	Doctor of Philosophy
YEAR THIS DEGREE GRANTED:	Spring, 1999

Permission is hereby granted to the University of Alberta Library to reproduce single copies of this thesis and to lend or sell such copies for private, scholarly or scientific research purposes only.

The author reserves all other publication and other rights in association with the copyright in the thesis, and except as hereinbefore provided neither the thesis nor any substantial portion thereof may be printed or otherwise reproduced in any material form whatever without the author's prior written permission.



10610 83rd Avenue #108
Edmonton, Alberta
CANADA
T6E 2E2

Date: Nov. 4/98

UNIVERSITY OF ALBERTA

FACULTY OF GRADUATE STUDIES AND RESEARCH

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research for acceptance, a thesis entitled ASPARTIC PROTEINASES: STRUCTURE, INHIBITION, AND ZYMOGEN ACTIVATION submitted by AMIR RAFIQ KHAN in partial fulfillment of the requirement for the degree of DOCTOR OF PHILOSOPHY.

Dr. Michael N.G. James

Dr. Charles F.B. Holmes

Dr. Robert O. Ryan

Dr. Brian D. Sykes

Dr. John C. Vederas

Dr. David R. Davies

Date: Oct 28/98

For my parents.

ABSTRACT

The catalytic activity of the aspartic proteinases, and proteolytic enzymes in general, must be tightly controlled by cells to prevent unwanted protein degradation. The human digestive enzymes, pepsins and gastricsins, are synthesized as inactive precursors (or 'zymogens') with an inhibitory N-terminal extension, or 'prosegment' (1p-43p; the suffix "p" denotes the prosegment), by mucosal cells that line the stomach. The zymogens are secreted into the low pH environment of the gastric juice where the prosegment is removed by an autocatalytic and stepwise mechanism to generate the ~320-residue mature enzyme. In these studies, an activation intermediate of human gastricsin that is transiently formed during the conversion pathway has been crystallized and its structure determined. The structure reveals that a part of the prosegment, residues 27p-43p, has been cleaved and removed. Large conformational changes are observed in loops adjacent to the active site, and the overall studies suggest an activation pathway that may be common to all gastric aspartic proteinases.

Once proteolytic enzymes are activated, their prosegments are generally poor inhibitors. However, interest in the development of inhibitors against aspartic proteinases has been sparked by their role in a variety of human pathogens. In addition to the optimization of chemical interactions between the enzyme and inhibitor, the binding affinity can be increased by constraining the inhibitor to the conformation that is recognized by the enzyme, thus lowering the entropic barrier to complex formation. A new phosphonate-based peptide inhibitor (PP7; K_i = 0.10 nM) was previously synthesized against penicillopepsin, a fungal aspartic proteinase. The inhibitor was designed with a covalent bridge between the P2 and P1' sidechains (nomenclature of Schechter & Berger, 1967) in order to reduce the conformational flexibility of the inhibitor and thereby enhance the binding affinity. This thesis will present the crystal structures of the P2-P1' macrocyclic inhibitor, along with a control acyclic compound (PP8; K_i = 42 nM), in complexes with penicillopepsin. Comparisons of the two crystal structures reveal that the conformations of the inhibitors and their interactions with the enzyme are similar. The 420-fold increase in the binding affinity of PP7 over PP8 is attributed to a reduction in its conformational flexibility, thus providing the first rigorous measure of the entropic contribution to the binding energy in a protein-ligand complex.

ACKNOWLEDGMENTS

I would like to thank Michael James for his guidance and support during my PhD studies. The last four years in Edmonton have been an exciting and rewarding time period, and without Michael's enthusiasm and wealth of knowledge, it would not have been possible.

Everyone in the lab has provided friendship and support during my studies. I'd like to thank Marie Fraser for her patience and advice, as well as Ernst, Masao, Nina, Kathy, Steve, Jens, Maia, and the rest of the lab for their help regarding all aspects of structure determination. Randy Read and members of his lab were a valuable resource for many aspects of crystallographic refinement.

Mae, Sue, and Gail were always willing to help me in the office with regard to bureaucratic burdens, manuscripts, and correspondences. Also, I thank every member of the Friday evening 'Avenue Pizza' club for making my Edmonton stay an enjoyable experience, even in the midst of winter.

Our collaborators at NIH (Dr. Nadezhda Tarasova), the University of California at Berkeley (Dr. Paul Bartlett), and Rice University (Dr. Janet Braam) were always eager to share ideas, and have provided outstanding support for our research projects.

Finally, I wish to aknowledge the financial support of the Medical Research Council of Canada and the University of Alberta (through a Dissertation Fellowship).

TABLE OF CONTENTS

CHAPTER 1: General Introduction

A. Biological Functions and Historical Perspective of the Aspartic Proteinases..... 1

B. Structure and Catalytic Mechanism of Aspartic Proteinases..... 6

C. Aspartic Proteinase Zymogens and their Activation Pathways..... 16

D. Research Goals..... 22

E. References..... 24

CHAPTER 2: Structures of cyclic and acyclic peptide-based phosphonate inhibitors in complexes with penicillopepsin

A. Introduction..... 35

B. Materials and Methods..... 47

C. Results..... 53

D. Discussion..... 71

E. References..... 75

CHAPTER 3: Structural Characterization of Activation 'Intermediate 2' on the Pathway to Human Gastricsin

A. Introduction..... 79

B. Materials and Methods..... 87

C. Results and Discussion..... 90

D. References..... 106

CHAPTER 4: General Discussion and Conclusions

A. Ultra-high Resolution Crystal Structure of PP7-Penicillopepsin..... 111

B. Conversion of Zymogens to Active Proteolytic Enzymes..... 126

C. Future Directions..... 128

D. References..... 132

LIST OF TABLES

Table 1A.1 Biological functions of the aspartic proteinases..... 3

Table 2A.1 Structures of penicillopepsin-inhibitor complexes..... 42

Table 2B.1 Crystal data for PP7 and PP8 complexes with penicillopepsin..... 49

Table 2B.2 Stages in the refinement of PP7-penicillopepsin..... 51

Table 2C.1 Refinement statistics for the penicillopepsin-phosphonate inhibitor complexes..... 54

Table 2C.2 Possible hydrogen-bonding contact distances between phosphonate inhibitors and the active site..... 66

Table 2C.3 Solvent-accessible surface area..... 67

Table 2C.4 Conformational angles of phosphonate inhibitors..... 68

Table 2C.5 Least-squares superpositions of penicillopepsin and the inhibitors in various complexes..... 69

Table 3B.1 Statsitics of the data and refinement..... 88

Table 4A.1 High-resolution crystal structures of proteins..... 112

Table 4A.2 Statistics of the ω-angle and its distortion from planarity in penicillopepsin structures..... 120

LIST OF FIGURES

Figure 1A.1 A sketch of the mucosal wall lining the stomach..... 4

Figure 1B.1 Ribbon model of human pepsin..... 7

Figure 1B.2 Close-up of the active site of human pepsin..... 9

Figure 1B.3 Backside of the active site of human pepsin..... 10

Figure 1B.4 Catalytic mechanism of the aspartic proteinases..... 11

Figure 1B.5 Structure of the complex of pepstatin with pepsin..... 13

Figure 1C.1 Sequence alignements of the prosegments of gastric aspartic proteinases..... 20

Figure 1C.2 Ribbon model of human pepsinogen..... 21

Figure 1C.3 Close-up of the active site of human pepsinogen..... 22

Figure 2A.1 Chemical drawing of PP2, the pepstide-based phosphonate inhibitor of penicillopepsin..... 37

Figure 2A.2 Active site structure of the PP2-penicillopepsin complex.... 38

Figure 2A.3 Chemical structures of 'second-generation' phosphonate inhibitors of penicillopepsin..... 40

Figure 2A.4 Active site structure of the PP6 penicillopepsin complex and its acyclic control..... 43

Figure 2A.5 Chemical structures of the P2-P1' series of phosphonate inhibitors..... 46

Figure 2C.1 Stick model of PP8 bound to the active site of penicillopepsin, shown as ribbons..... 55

Figure 2C.2 Ball-and-stick models of PP7 and PP8 at the active site penicillopepsin..... 58

Figure 2C.3 Superpositions of PP7 and PP8 at the active site of penicillopepsin..... 59

Figure 2C.4 Solution NMR structure of PP7..... 62

Figure 2C.5 Chemical synthesis of PP7..... 63

Figure 2C.6 Evaluation of the rational design strategy..... 65

Figure 2C.7 Electron density at the active site of PP7-penicillopepsin.... 70

Figure 2D.1 Stick model of PP7 against the surface of penicillopepsin... 74

Figure 3A.1 Ribbon model of human progastricsin..... 80

Figure 3A.2 Superposition of human progastricsin with porcine pepsinogen, shown as C_α vectors..... 81

Figure 3A.3 Superposition of the prosegments of human progastricsin and porcine pepsinogen..... 82

Figure 3A.4 Analyses of the activation pathway for progastricsin by gel electrophoresis..... 83

Figure 3C.1 Ribbon model of hGSi, the activation 'intermediate 2' of human gastricsin..... 91

Figure 3C.2 Stick model of the prosegment β-strand of hGSi..... 92

Figure 3C.3 Ribbon model of the superposition of hGSi and progastricsin..... 94

Figure 3C.4 Elctron density at the active site of hGSi..... 95

Figure 3C.5 Ribbon model of the active site of progastricsin..... 98

Figure 3C.6 Ribbon model of the active site of hGSi..... 99

Figure 3C.7 Ribbon model of the active site of human pepsin..... 100

Figure 3C.8 Theoretical model of pepstatin at the active site of hGSi..... 102

Figure 3C.9 Coil rendering of the prosegments of pepsinogen and progastricsin against the enzyme surfaces..... 104

Figure 4A.1 A schematic of the short hydrogen bond between phosphate-binding protein and phosphate..... 114

Figure 4A.2 A schematic of the hydrogen bonding scenario at the active site of PP7-penicillopepsin..... 115

Figure 4A.3 A section of the Fo-Fc omit map of PP7-penicillopepsin contoured at 2.5σ..... 117

Figure 4A.4 Difference density at the active site of PP7-penicillopepsin..... 118

Figure 4A.5 Ball-and-stick model of PP7-penicillopepsin at the Gln306-Tyr307 peptide bond..... 121

Figure 4A.6 Analysis of disorder in PP7-penicillopepsin..... 125

Figure 4B.1 Crystallographic B-factors of the prosegments of various zymogens..... 127

LIST OF ABBREVIATIONS AND SYMBOLS

Å	angstroms (10^{-10} metres)
B	isotropic thermal motion parameter ($B=8\pi^2/3\langle r^2\rangle$)
CHESS	Cornell High Energy Synchrotron Source
DESY	Deutsches Elektronen-Synchrotron
esd	estimated standard deviation
|F|	diffracted X-ray structure factor amplitude
hPGC	human progastricsin
hPGN	human pepsinogen
I	diffracted X-ray scattering intensity
kDa	kilodaltons (1/12 mass of ^{12}C nucleus)
K_i	inhibition constant (moles/litre)
kV	electrical potential in kilo-volts
λ	wavelength
mA	electrical current in milli-amperes
NMR	nucelar magnetic resonance
p	suffix to indicate "prosegment", *e.g.* Lys37p
PPi	penicillopepsin phosphonate inhibitor
rms	root mean square
σ	standard deviation
ω-angle	backbone torsion angle $C_{\alpha,i}$-C_i-N_{i+1}-$C_{\alpha,i+1}$
χ2-angle	sidechain torsion angle C_α-C_β-C_γ-$C_{\delta 1}$

CHAPTER 1

INTRODUCTION

Proteolytic enzymes are found in organisms across all of the kingdoms of life. These enzymes catalyze the cleavage of peptide bonds in protein and peptide substrates, consuming water molecules in the process. Perhaps the most fundamental role of proteolytic enzymes is the non-specific digestion of other proteins to provide nutrients for living organisms. However, they also perform regulatory roles with narrow substrate specificities, and are implicated in such diverse biological processes as blood coagulation and fibrinolysis (Davie et et al., 1991), the immune response (Groettrup et al., 1996), tumour invasion (Edwards & Murphy, 1998), regulation of sodium balance and blood pressure (Vallet et al., 1997), cellular differentiation (Weinmaster, 1998), and programmed cell death (Greenburg, 1996).

Proteolytic enzymes are classified by the International Union of Biochemistry (IUBMB) according to their mechanism of peptide hydrolysis (IUBMB, 1992). The serine proteinases utilize a catalytic serine residue for the cleavage reaction; the metalloproteinases have a Zn^{2+} ion that is essential for catalysis; the cysteine proteinases have a catalytic cysteine residue at their active site; and finally, the aspartic proteinases have two catalytic aspartate residues that are involved in the reaction mechanism. This thesis will focus upon the aspartic proteinases for which the archetypal member is the digestive enzyme pepsin.

A. Biological Functions and Historical Perspective of the Aspartic Proteinases

The aspartic proteinases (EC 3.4.23.-) comprise a superfamily of endo-proteolytic enzymes that are found in a variety of organisms in nature. These enzymes perform essential roles in retroviruses, yeast, fungi, plants, and mammals.

The aspartic proteinases are named after the two catalytic Asp residues that form the active site, and their overall structure has been conserved across the evolutionary spectrum. Many members of this superfamily have optimal activity at low pH, hence they are also termed 'acid proteases' in the literature. Some of the biological roles of these enzymes include (1) site-specific proteolysis of the retroviral polyprotein precursor to generate the structural, regulatory, and enzymatic proteins that are necessary for replication of the virion (e.g., HIV-1 proteinase, Wlodawer & Erickson, 1993); (2) a role in tissue invasion and virulence for fungal pathogens of humans (Cutfield et al., 1995); (3) cleavage of hemoglobin in the digestive vacuoles of the malarial parasite (plasmepsins, Francis et al., 1997); and (4) regulation of blood pressure (renin, Sielecki et al., 1989; Table 1A.1). The aspartic proteinases also have a long history in the field of human cuisine. The enzyme chymosin has been an active ingredient in cheese-making for thousands of years, and aspartic proteinases are also used for the production of soya sauce, a practice that apparently originated during the Zhou dynasty (Hofmann, 1989; Davies, 1990).

Pepsins and gastricsins. Perhaps the most basic role of aspartic proteinases, and proteolytic enzymes in general, is the digestion and catabolism of other proteins. Mammals secrete the 'gastric' aspartic proteinases into the low pH environment of the stomach, where they function as digestive enzymes that degrade dietary proteins. Most proteins are denatured in the stomach (pH < 2) and are susceptible to digestion by 'pepsin' and 'gastricsin', a pair of aspartic proteinases that function optimally at low pH (Fig. 1A.1). Pepsin is the archetypal aspartic proteinase, and was originally named in the 1830s by Schwann after the Greek word *pepsis*, meaning 'digestion'. A century later, porcine pepsin figured prominently during the early attempts to characterize the structure of proteins. The enzyme was the first protein for which the diffraction of X-rays was demonstrated in the crystalline state (Bernal & Crowfoot,

Table 1A.1: Biological functions of the Aspartic Proteinases

Sub-family	Physiological location	Biological function
Mammalian gastric enzymes	stomach	digestion of dietary proteins
fungal enzymes	secreted from fungi (extracellular)	tissue invasion
renin	secreted into blood	cleaves angio-tensinogen during blood co-agulation
retroviral enzymes	HIV and other retroviruses	cleavage of viral polyprotein precursor
cathepsin D	lysosomes	protein degradation antigen presentation
plasmepsin	digestive vacuoles of malarial parasite	cleavage of hemoglobin

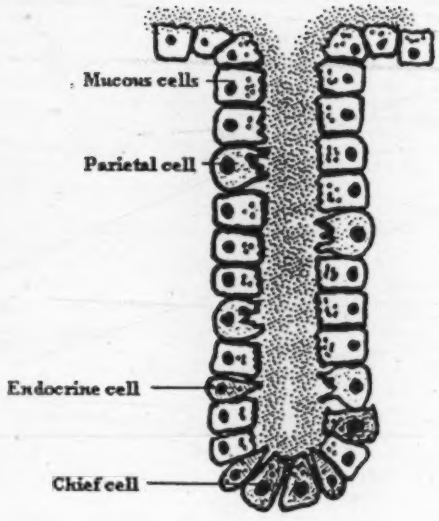


Figure 1A.1: A sketch of the mucosal wall lining the stomach. Acid is secreted by the parietal cells. The mucosal and chief cells secrete pepsinogens and progastricsin, respectively, which are the precursors of pepsins and gastricsin. The figure was taken from Fusek & Vetvicka (1995).

1934), suggesting that the elusive structures of proteins might be determined using crystallographic diffraction techniques.

Biology of penicillopepsin and other fungal enzymes. Extracellular and intracellular proteinases have been implicated as virulence factors for several fungi and yeast-like fungal pathogens that infect mammals (Kothary et al., 1984; Kwon-chung et al., 1985). For example, the genus *Candida* comprises a family of yeast-like fungi that are a major cause of infections in immunocompromised humans (Fusek & Vetvicka, 1995). Although the precise role of aspartic proteinases remains ambiguous, it is suspected that these enzymes may provide metabolic nutrients to the growing fungus (Thangamani & Hofmann, 1965). In addition, proteolytic activity may be important for degrading mucosal surfaces during the invasion process (Monod et al., 1994). Although the relative importance of aspartic proteinases during the various stages of parasitic infections remains a matter of dispute (Germaine & Tellefson, 1981; Togni et al., 1994), there is general agreement that they are associated with higher aggressiveness in human infections.

The secreted aspartic proteinase from *Penicillium janthinellum* is an evolutionary homologue of the gastric aspartic proteinases, sharing about 30% sequence identity and a similar three-dimensional structure (Sodek & Hofmann, 1970; Hsu et al., 1977). The 323-residue enzyme 'penicillopepsin' was originally identified by Kunitz (1938), who speculated that the proteinase may directly induce sporulation since its appearance coincides with this process. However, later biochemical studies showed that there is no direct relationship between penicillopepsin activity and sporulation (Thangamani & Hofmann, 1966). Penicillopepsin was later among the first aspartic proteinases to be characterized by X-ray crystallography (Hsu et al., 1977; James & Sielecki, 1983). In recent years, it was noticed that the crystals of penicillopepsin are of exceptional quality, diffracting to about 1.3 Å using laboratory

X-ray generators (Ding et al., 1998). Modern day crystallographic resources at the synchrotron, including sensitive and rapid data acquisition instruments and high intensity X-ray sources, have now permitted data collection and characterization of penicillopepsin to beyond 1.0 Å resolution (Chapter 2).

In summary, aspartic proteinases are essential for diverse biological processes. They are relevant to human diseases such as Acquired Immunodeficiency Syndrome (AIDS), malaria, and fungal infections. Also, they have commercial importance to the food industry, for example, in the fermentation processes that generate cheese and soya sauce. This thesis will address several questions related to the structure, mechanism and inhibition of the aspartic proteinases using penicillopepsin and the mammalian gastric proteinases (pepsin, gastricsin) as model enzymes.

B. Structure and Catalytic Mechanism of Aspartic Proteinases

Aspartic proteinases have a bilobate shape resembling a 'croissant' or 'kidney'. They consist of two β-barrel domains that are connected by a central six-stranded β-sheet. A long hydrophobic groove between the two domains forms the substrate binding cleft (Fig. 1B.1). The two catalytic aspartate residues (Asp32 and Asp215 in pepsin numbering) reside on two similarly-folded loops at the base of the hydrophobic groove, in front of the central β-sheet. There is a pseudo two-fold symmetry in human pepsin that is also reflected in the approximate two-fold symmetry at the active site residues and the loops upon which the Asp32 and Asp215 reside. Upon division of the two lobes of penicillopepsin at the central β-sheet, they can be superimposed with a root-mean-square (rms) deviation of 1.91 Å for 70 equivalent C_α atoms (James & Sielecki, 1983). The viral aspartic proteinases have evolved as a compact single β-barrel domain that dimerizes to form the active enzyme (Miller et al., 1989; Navia et al., 1989). Each monomer contributes one of the two

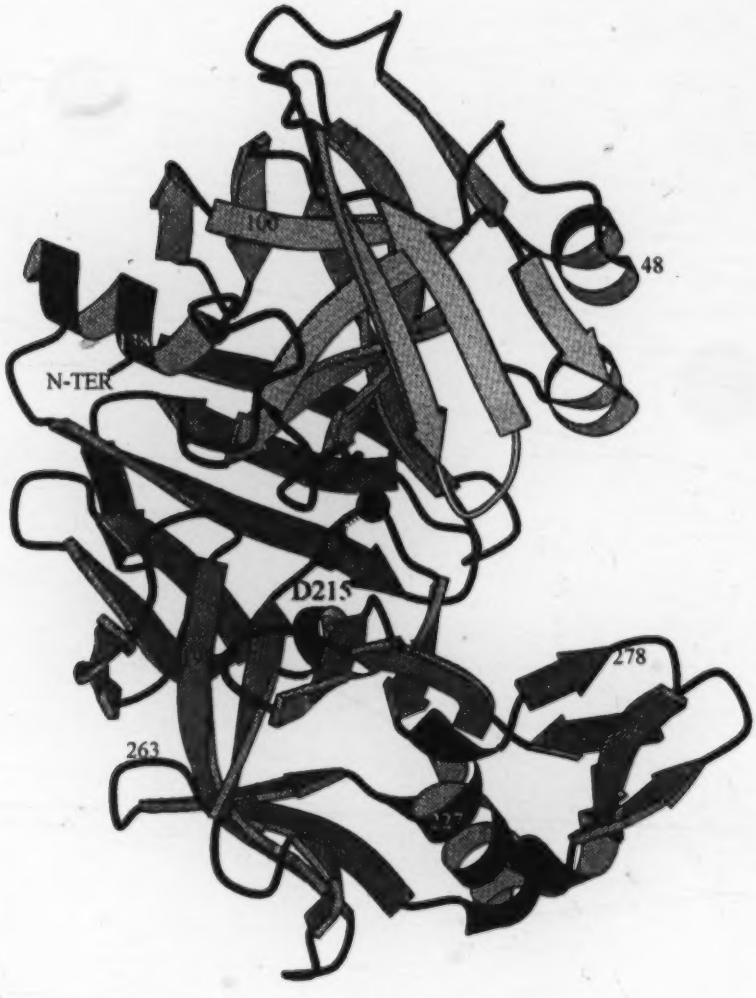


Figure 1B.1: Ribbon model of human pepsin (Fujinaga et al., 1995). The various β-sheets in the structure are colour coded. The N-terminal β-barrel is yellow; the central six-stranded β-sheet is purple; and the C-terminal β-barrel, which is topologically equivalent to the N-terminal β-sheet, is blue. The nucleophilic water molecule is represented by a sphere. The N-terminal β-strand (residues 2-8) is coloured green. Asp32 and Asp215 are represented as red stick models. The figures in this chapter were drawn using BOBSCRIPT (Esnouf, 1997).

aspartates which comprise the active site of the dimer, and this is consistent with the prediction that cellular aspartic proteinases have emerged by gene duplication of a single domain (Tang et al., 1978).

There is consensus that peptide hydrolysis by aspartic proteinases involves a non-covalent general acid-base mechanism, as was originally proposed by Fruton (1976). This hypothesis was supported by the inability to isolate covalent intermediates in the reaction (Dunn & Fink, 1984; Fruton, 1976; Hofmann et al., 1984), and by isotope exchange experiments with ^{18}O water (Antonov et al., 1981). The structures of all active aspartic proteinases contain a water molecule that is hydrogen-bonded to the two catalytic Asp residues (Fig. 1B.2,3). In the proposed hydrolytic pathway, this central water is de-protonated and subsequently attacks the carbonyl carbon of the scissile bond of substrates to form a tetrahedral intermediate (Suguna et al., 1987; Davies, 1990; James et al., 1992).

Since the location of hydrogens cannot be determined at the resolution of the structures of aspartic proteinases that have been characterized, the detailed pathway for the reaction remains a matter of debate. However, one model of catalysis considers Asp215 to have a negative charge, whereas Asp32 is protonated prior to the reaction (Suguna et al., 1987; James et al., 1992). During the first step in hydrolysis, Asp215 acts as the general base, extracting a proton from the nucleophilic water molecule (Fig. 1B.4). The water proceeds to attack the carbonyl carbon of the scissile bond, thereby forming a tetrahedral intermediate. The nucleophilic attack and subsequent breakdown of the tetrahedral intermediate is assisted by the side chain of Asp32, which acts as the general acid and forms hydrogen bonds with the negatively-charged oxyanion carbonyl oxygen. Furthermore, there are additional hydrogen bonding interactions between the catalytic Asp residues and neighbouring amino acids that likely influence the distribution and orientation of protons and electrons at the active site (James et al., 1992).

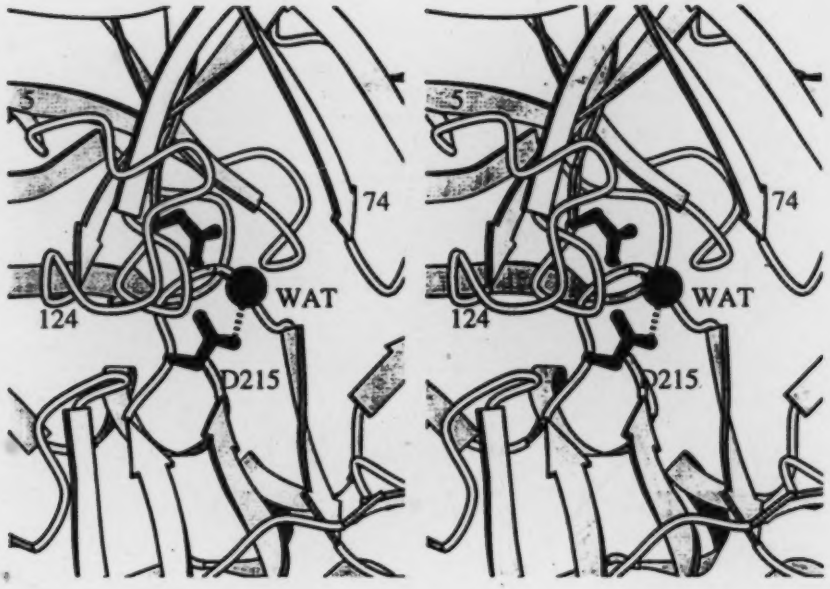


Figure 1B.2: Close-up of the active site of human pepsin, shown in divergent stereo. The catalytic aspartate residues (Asp32, top; Asp215, bottom) are shown as ball and stick models. The nucleophilic water molecule is represented by a sphere. Hydrogen bonds are shown as dashed lines.

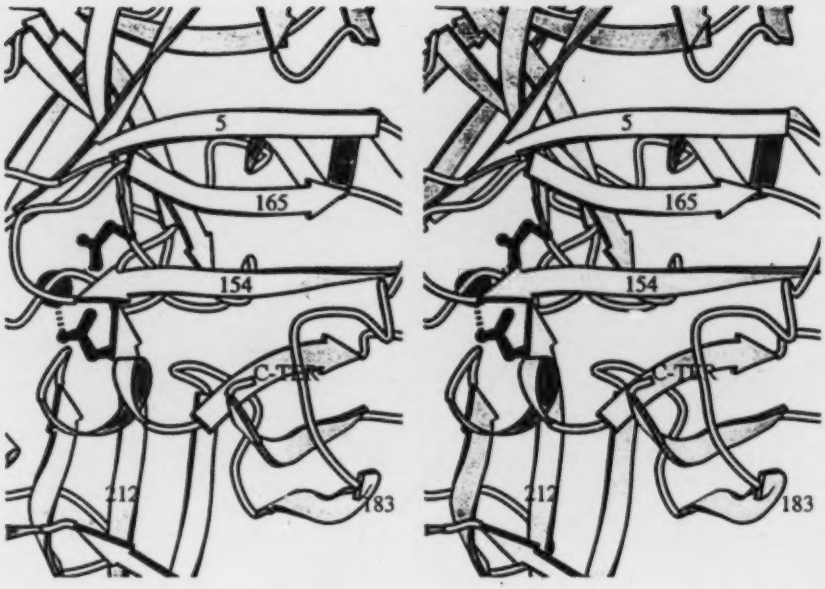


Figure 1B.3: Backside of the active site of human pepsin. The molecule has been rotated 180° around the vertical axis of the previous figure in order to provide a clear view of the central six-stranded β-sheet. The N-terminus of pepsin forms the first strand of this anti-parallel β-sheet, and the active is partially obscured on the opposite side. Asp32 is on top; Asp215 is on the bottom.

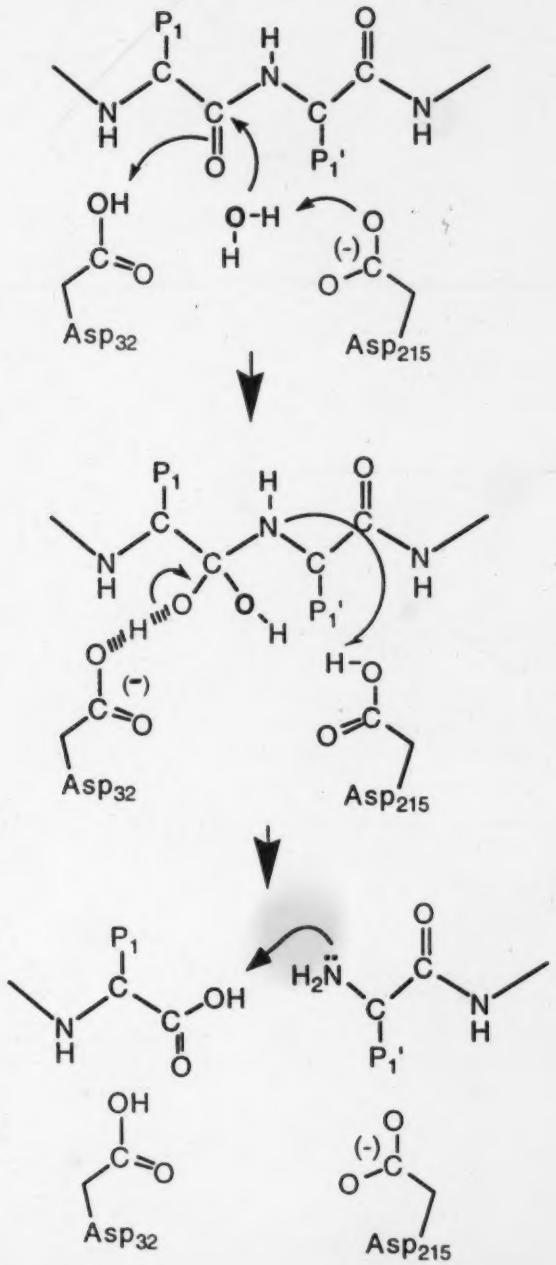


Figure 1B.4: Catalytic mechanism for the hydrolysis of peptide bonds by aspartic proteinases (adapted from Davies, 1990). The mechanism requires that one of the catalytic aspartates is charged, while the other is protonated.

Inhibitors of aspartic proteinases. Many peptide-based inhibitors have been characterized in complex with aspartic proteinases, thus providing clues to the mechanism for cleavage of substrates. For example, the peptide mimic 'pepstatin' from *Streptomyces* filtrates is a potent inhibitor of most aspartic proteinases, supporting the view that these enzymes share a common catalytic mechanism. Pepstatin contains an unusual 'statine' moiety [(3S,4S)-4-amino-3-hydroxy-6-methylheptanoic acid] at the P1 position (notation of Schechter & Berger, 1967; see below, *Substrate binding and specificity*). The *3S*-OH group in the backbone of statine replaces the nucleophilic water at the active site, and the overall complex resembles the tetrahedral intermediate of the hydrolytic pathway of a good substrate (Davies, 1990; Fig. 1B.5).

Synthetic phosphorus-containing analogues of peptides have also been analyzed in complex with aspartic proteinases. Originially, phosphorus-containing peptide mimics were observed to be excellent inhibitors of another class of enzymes, the zinc metalloproteinases. Characterization of their interactions with the zinc enzymes revealed that some of these inhibitors mimic the cleavage transition state (Weaver et al., 1977; Bartlett & Marlowe, 1983; Tronrud et al., 1987). Since the zinc enzymes cleave peptide bonds in an analogous manner to the aspartic enzymes, namely *via* the activation of a water nucleophile, phosphorus-containing oligopeptides were synthesized and tested against several aspartic proteinases (Bartlett & Kazer, 1984; Bartlett et al., 1990). These inhibition studies proved that phosphorus-based peptides are indeed inhibitors of enzymes such as pepsin (Bartlett & Kazer, 1984), renin (Allen et al., 1989), HIV protease (Dreyer et al., 1989) and penicillopepsin (Bartlett et al., 1990). Subsequently, our lab crystallized penicillopepsin in complex with several of these inhibitors in order to characterize the inhibitory interactions and to provide insight into the catalytic mechanism (Fraser et al., 1992). In Chapter 2, these studies will be extended to novel "second-generation" inhibitors of

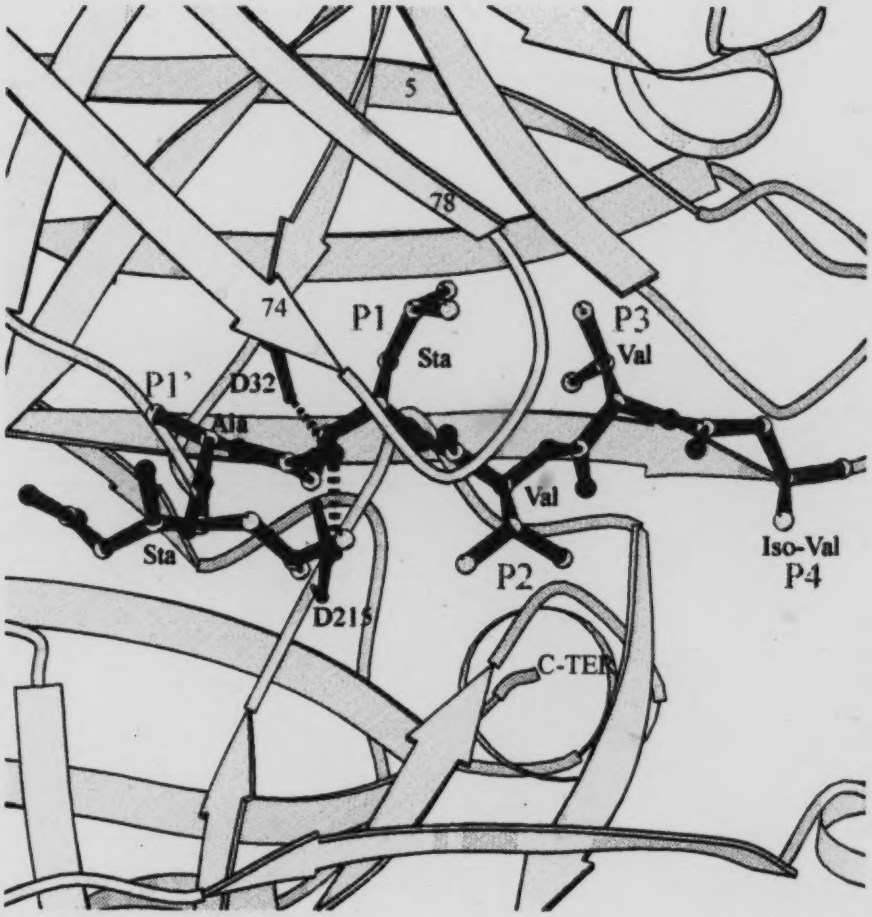


Figure 1B.5: Structure of the complex of pepsin with pepstatin (Fujinaga et al., 1995). The inhibitor resides at the active site in an extended conformation (the N-terminus of the inhibitor is at the right, beginning with Iso-Val). The (*S*)-3-hydroxyl of the P1-statine residue hydrogen bonds to Asp32 and Asp215 (shown as dashed lines).

penicillopepsin that have been designed to address several questions arising from the initial work by Fraser et al. (1992).

Substrate-binding and specificity. The typical substrate-binding cleft of aspartic proteinases is sufficiently long to accommodate a minimum of 6 amino acids, such that four residues extend towards the N-terminal side of the cleaved peptide bond (P1→P4) and two residues are present on the C-terminal side (P1'→P2'; nomenclature of Schechter & Berger, 1967). The corresponding enzyme specificity pockets that interact with these residues are S1→S4 and S1'→S2'. Crystallographic studies of enzyme complexes with pepstatin and other peptide-based inhibitors have revealed that substrates bind to the active site in an extended conformation (Fig. 1B.5). The extended substrate-binding cleft allows selective cleavage of target peptides, and the effect of strain energy in twisting the P1-P1' scissile bond away from its normal planar conformation may also contribute to the cleavage reaction (Fruton, 1976; Davies, 1990). In support of this notion, elongation of peptide substrates (from four to six amino acids) on either side of the peptide bond increases k_{cat} by several orders of magnitude, while having limited effects upon the K_m (Fruton, 1976).

Extensive kinetic experiments using peptide substrates and inhibitors have characterized the substrate-binding pockets and specificities of these enzymes. Many aspartic proteinases have a preference for large hydrophobic residues flanking the scissile peptide bond, at the P1 and P1' positions. The structural basis for this preference is the predominantly hydrophobic nature of the corresponding S1 and S1' specificity pockets. Based upon these obervations, Ben Dunn's group synthesized a library of peptide inhibitors that were built around a core of Phe-NPhe at the P1-P1' positions (Nphe = *p*-nitro-phenylalanine). These inhibitors were designed to probe the differences in specificity of the aspartic proteinases at binding pockets that are

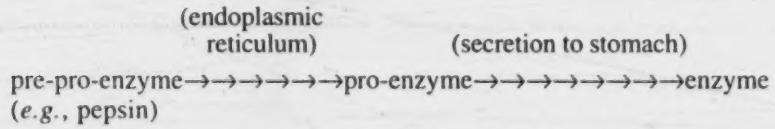
distal from the scissile bond (Dunn et al., 1998; Rao & Dunn, 1995; Dunn et al., 1994). It was concluded that the 'secondary specificity' of the enzymes may have a significant impact upon the kinetics of cleavage, and could be exploited to develop specific inhibitors of biologically important enzymes. However, not all aspartic proteinases have a preference for hydrophobic residues at the P1 and P1' positions. Mammalian 'yapsin', a newly-characterized aspartic proteinase that specifically cleaves peptide pro-hormones to their active counterparts, cleaves with a positively-charged residue at the P1 position (Olsen et al., 1998). Comparative modeling of the three-dimensional structure of yapsin predicted a cluster of negatively charged residues at the S1, S2' and S3' sites that likely interact with positively charged substrate residues at the P1, P2' and P3' positions (Olsen et al., 1998).

Finally, in addition to the substrate-binding cleft, aspartic proteinases contain a β-hairpin loop that hangs over the active site, and it is often termed the "flap" in the literature (Fig. 1B.5; residues 70-80 of pepsin). The tip of the flap in the cellularly-derived enzymes contains an invariable Tyr residue (Tyr75 in pepsin), and the flap is flexible in the native (uncomplexed) structures of the enzymes, as evidenced by the high *B*-factors of the backbone atoms. However, in the structures of aspartic proteinase complexes, the flap clamps down upon substrates and becomes ordered, forming van der Waals and hydrogen bond interactions with segments of the peptide (*e.g.*, human pepsin, Fujinaga et al., 1995; penicillopepsin, James & Sielecki, 1985). In summary, the aspartic proteinase structures contain an extended substrate-binding cleft that is generally hydrophobic and accommodates peptides with an approximate β-strand-like conformation. The precise substrate specificities reflect the composition of the various binding subsites on the enzyme, and interactions that are distant from the scissile peptide bond may have a significant influence upon the efficiency of cleavage. For example, hydrogen bonds between the enzyme and subsites S2' and S3 may facilitate the distortion of the scissile pepstide bond toward a productive enzyme-

substrate complex (Hofmann et al., 1988; Pearl, 1987; Suguna et al., 1987; Fruton, 1976).

C. Aspartic Proteinase Zymogens and their Activation Pathways

Aspartic proteinases are initially synthesized within cells as inactive precursors (or 'zymogens'), with an inhibitory N-terminal extension (or 'prosegment'). In fact, the prosegments of cellular enzymes have an additional extension, the 'pre-sequence' (or signal sequence), which targets the enzyme precursor to the secretory pathway. The activation process of the precursor enzyme requires successive removal of the signal sequence within the secretory pathway, followed by limited proteolysis and removal of the prosegment within the appropriate biological compartment, thus generating the active enzyme. Many other classes of proteolytic enzymes are also synthesized as zymogens, with an inhibitory N-terminal prosegment (*e.g.*, serine proteinases, cysteine proteinases, metalloproteinases; reviewed by Khan & James, 1998). The broad function of the prosegment is to prevent proteolysis until the zymogen is targetted to the appropriate compartment in which the active enzyme exerts its biological activity. However, the prosegments of aspartic proteinases may also serve other functions, such as providing a molecular recognition motif during intracellular sorting (yeast pro-proteinase A, Klionsky et al., 1988; Westphal et al., 1996). Since Chapter 3 will address the activation pathway of the gastric aspartic proteinases, this sub-family of enzymes will be the focus of the ensuing discussion.



At this point, it is important to clarify the terms 'conversion' and 'activation'. Although these terms will be used interchangeably, it is essential to point out that the

zymogen forms of gastric aspartic proteinases are "active" in the sense that the catalytic machinery (referring to the active site aspartates and substrate cleft) is preformed and is capable of hydrolyzing peptide bonds. Indeed, this is a prerequisite for the autocatalytic mechanism leading from the zymogen to the mature enzyme. Therefore, the terms 'activation' and 'conversion' refer to the events that result in the cleavage and removal of the inhibitory prosegment and formation of the mature enzyme.

Activation studies of the gastric enzymes. The mammalian gastric aspartic proteinases, gastricsins and pepsins (which share 50% sequence identity), are synthesized in the chief cells of the stomach as zymogens, and contain an inhibitory N-terminal prosegment of 43 to 47 residues in length. The mature segments of these enzymes are between 320 and 330 residues in length. Conversion to the active enzymes occurs by limited proteolysis and removal of the prosegment upon secretion of the zymogens into the low pH (about 1.5) environment of the gastric juice. These enzymes have optimal activity at low pH, and digest dietary proteins that become acid-denatured in the stomach. The table below summarizes the frequencies of charged residues in the zymogens, showing clearly the enrichment of positively-charged amino acids in the prosegments and the preponderance of negatively-charged residues in the mature segment. As will be discussed shortly, salt bridges between the prosegment and the mature segment are essential for stabilizing the zymogen at neutral pH, and it is the weakening of these interactions at low pH that initiates the conversion to the active enzyme.

	Prosegment		Mature Segment	
	Asp/Glu	Arg/Lys	Asp/Glu	Arg/Lys
pepsinogen	3/44	11/44	42/326	3/326
progastricsin	5/43	11/43	26/329	3/329

The overall distribution of pepsinogen *vs.* progastricsin in the stomach is different; pepsinogen A is localized mainly to the stomach, whereas progastricsin is found throughout the stomach (Samloff & Liebmann, 1973). Also, human pepsinogens can be separated into at least five distinct bands by polacrylamide gel electrophoresis (Samloff, 1969). The heterogeneity is attributed to multiple copies of the zymogens ('isozymogens') that differ slightly in amino acid sequences, and to differences in post-translational modifications (Defize et al., 1985; Taggart et al., 1985). In one case, a single amino acid substitution in the prosegment (Lys43p to Glu43p) confers the ability to resolve the isozymogens by electrophoresis (Athauda et al., 1989). The biological significance of the distinct location of pepsin in the stomach, as well as the genetic polymorphism of human pepsinogens, is not understood.

The activation process of the aspartic proteinases has been extensively studied using biochemical and biophysical techniques, and has a rich place in the history of biochemistry. During the late 1930s, Roger Herriott analyzed the kinetics of the conversion of porcine pepsinogen to pepsin at low pH and observed a first-order rate constant during the initial events (Herriott, 1938/1939). Furthermore, he identified an intermediate compound, noting that the initial loss of pepsinogen was not immediately accompanied by an increase in pepsin (Herriott, 1939). These classical experiments revealed that the conversion of pepsinogen→pepsin at low pH is an autocatalytic

reaction, and that the pathway involves the formation of transient intermediate(s), pepsinogen→[intermediate]$_n$→pepsin.

Structural studies of zymogens. Whereas the structures of the active enzymes are known from many organisms, the only zymogen structures that have been determined are from mammalian sources. These zymogens are porcine pepsinogen (pPGN), human pepsinogen A (hPGA), and human progastricsin (hPGC), which are all components of the gastric juices of the respective species (James & Sielecki, 1986; Bateman et al., 1998; Moore et al., 1995). The zymogen forms of gastric proteinases contain a positively-charged N-terminal prosegment that wraps around the central portion of the enzyme, forming salt-bridge interactions with the negatively-charged mature segment. The prosegments of the gastric zymogens vary between 43 to 47 residues in length and are related in both sequence and structure (Fig. 1C.1). The prosegment is not an independently-folding unit, and once removed, the peptide is autocatalytically degraded by the active enzyme.

The structure of porcine pepsinogen, the first published structure of a zymogen (James & Sielecki, 1986), revealed that a Lys residue (Lys36p in pPGN; "p" suffix denotes the prosegment) forms hydrogen-bonded salt bridges to the active site Asp residues, Asp32 and Asp215 (Figs. 1C.2,3). These interactions position the prosegment region Pro33p-Pro39p (a 3_{10}-helix) in front of the preformed active site, thereby preventing the approach of substrates. The conversion process is initiated by disruption of these salt bridges at low pH and subsequent limited proteolysis of the prosegment, as originally proposed by Bent Foltmann (1981). No other accessory molecules are required for conversion. Comparisons of the structures of the zymogens (pepsinogen and progastricsin) with active enzymes have shown that upon proteolytic removal of the prosegment, the N-terminus (approximately residues 1-10) of the mature segment undergoes a major conformational re-arrangement. In the

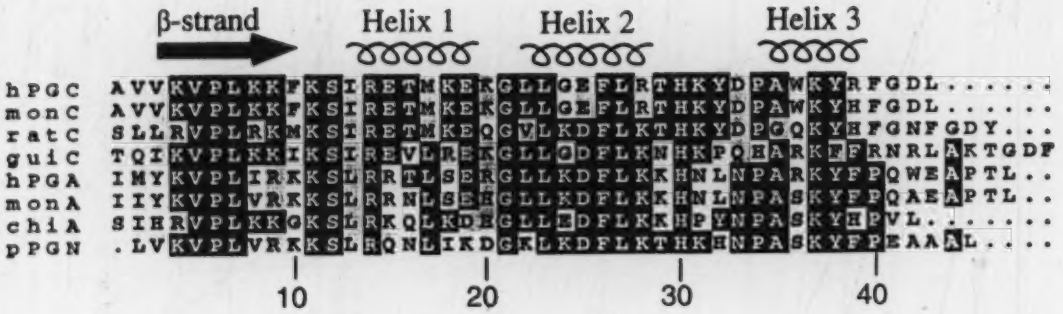


Figure 1C.1: Sequence alignments of the prosegments of gastric aspartic proteinases. The sequences correspond to human progastricsin (hPGC; Hayano et al., 1988); monkey progastricsin (monC; Kageyama & Takahashi, 1986a); rat progastricsin (ratC; Ichihara et al., 1986); guinea pig progastricsin (guiC; Kageyama et al., 1992); human pepsinogen (hPGA; Sogawa et al, 1983); monkey pepsinogen (monA; Kageyama & Takahashi, 1986b); chicken pepsinogen (chiA; Baudys & Kostka, 1983); and pig pepsinogen (pPGN; Foltmann, 1988). The secondary structure assignments are based upon the crystal structures of pig pepsinogen and human progastricsin.

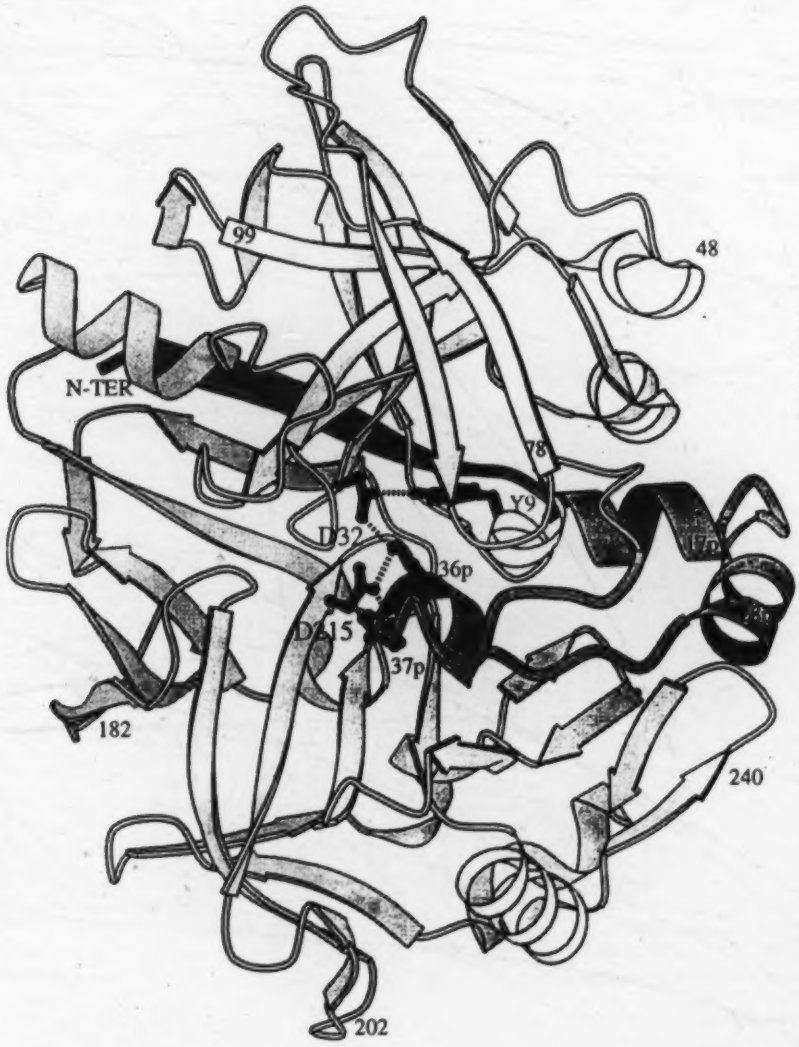


Figure 1C.2: Ribbon model of human pepsinogen. The prosegment is highlighted by the darker shading. Hydrogen bonds and salt bridges are represented as dashed lines. The prosegment contains three helices (α1p-α3p), and it is Ly36p from α3p that forms salt bridges with the active site aspartate residues and prevents access to substrates. The N-terminus of the prosegment forms the first strand of the central six-stranded β-sheet, a position that will eventually become occupied by the N-terminus of the mature segment in the active enzyme.

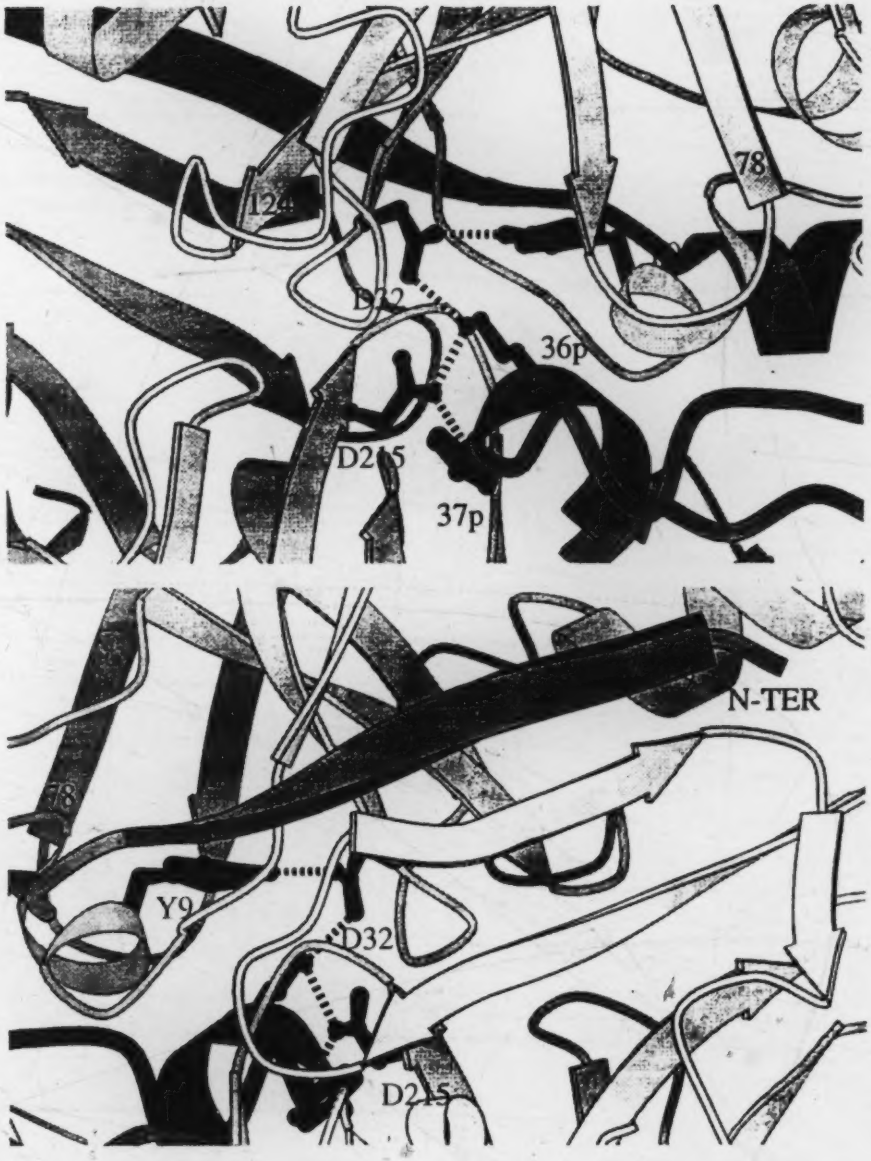


Figure 1C.3: Close-up of the active site of human pepsinogen, looking directly into the active site (top) and from behind (bottom). The prosegment is emphasized in darker shading, as in the previous figure. Tyr37p and Tyr9 form hydrogen bonds with Asp215 and Asp33, respectively, in a pseudo two-fold symmetric fashion.

zymogen, this portion of the mature segment consists of a loop and 3_{10} helix that protrudes into the active site cleft. However, in the mature enzyme, residues Ser1-Met10 swing around to the back of the molecule, forming the first strand of the six-stranded β-sheet. In the zymogen, prosegment residues Ala1p to Lys10p comprise the equivalent β-strand (Fig. 1C.3), but the prosegment is removed during the conversion process.

The early steps in the pathway leading to these remarkable structural changes have also been characterized indirectly using spectroscopic techniques in the 1970s and 1980s (McPhie, 1972; Auer & Glick, 1984). Upon initial exposure of the zymogen to acid, a conformational change occurs without proteolytic cleavage in the 5ms to 2s timescale (Auer & Glick, 1984). These studies involved stopped-flow fluorescence experiments using intrinsic and extrinsic fluorescent-tagged porcine pepsinogen. The conformational changes were inferred from fluorescence changes upon incubation of pepsinogen at pH < 3, and the timescale attributed to these changes was an order of magnitude faster than the dissociation of the prosegment from the mature segment (Twining et al., 1981). In addition, intermediates that form upon partial cleavage of the prosegment at much longer time scales have been characterized using various biochemical techniques in the labs of Bent Foltmann in Copenhagen (Nielsen & Foltmann, 1993), Kenji Takahashi in Japan (Kageyama et al., 1989), and Jordan Tang in Oklahoma (Marciniszyn et al., 1976). In Chapter 3, the first X-ray crystallographic structure of an intermediate aspartic proteinase ('intermediate 2' of human gastricsin) will be presented. The structure has properties resembling both the zymogen and the active enzyme, and has broad implications for the nature and biological significance of the conversion process.

D. Research Goals

The broad aim of this research project is further structural and functional characterization of aspartic proteinases and their zymogens. The specific goals are listed under the corresponding Chapters, below:

Chapter 2. The structures of penicillopepsin in complex with two newly-synthesized phosphonate inhibitors will be described. One of the inhibitors has been bridged at the side chains corresponding to the P2 and P1' positions in order to reduce the conformational flexibility of the peptide. Structural and kinetic comparisons of this "macrocyclic" inhibitor with its acyclic analogue will provide an estimation of the entropic cost of binding a flexible peptide to the enzyme active site.

Chapter 3. The structure of activation 'intermediate 2' on the pathway to human gastricsin will be presented. The structure of the intermediate will complement the previously determined structures of zymogens (human progastricsin, human pepsinogen) and the active enzymes (*e.g.*, human pepsin), thus providing molecular details of various stages in the activation pathway of gastric aspartic proteinases. Analyses of these structures will enable an understanding of the biological significance of the conversion process.

Chapter 4. General Discussions. A section will be devoted to inhibitory mechanisms that are employed in the context of zymogens *vs.* the active enzymes. In many cases, the active sites of both the zymogen and the active enzyme are competent for performing the cleavage reaction. For example, the catalytic aspartate residues in gastric zymogens are observed in the active conformation and position that is necessary for catalysis. However, the inhibitory interactions between the prosegment and the active site (Chapter 3) are distinct from the mechanism utilized by peptide-

based competitive inhibitors (Chapter 2) that bind tightly and often mimic the cleavage transition state.

A brief section of Chapter 4 will touch upon the importance of the high resolution crystal structure of PP7-penicillopepsin. The structure is the highest resolution that has been obtained for an aspartic proteinase, and the ~3000 atoms in the model represent one of the largest structures that has been determined in the protein database. Consequently, the structure provides a wealth of information that will require detailed analyses of the conformation, geometry, bond lengths/angles, and disorder. However, such in-depth analyses will not be presented in this thesis. This work is currently being performed in concert with studies of native penicillopepsin and other inhibitor-penicillopepsin complexes at atomic resolution (Fraser et al., unpublished work).

E. REFERENCES

Allen, M.C., Fuhrer, W., Tuck, B., Wade, R. & Wood, J.M. (1989) Renin inhibitors. Synthesis of transition-state analogue inhibitors containing phosphorus acid derivatives at the scissile bond. *J Med Chem* **32**: 1652-1661.

Antonov, V.K., Ginodman, L.M., Rumsh, L.D., Kapitannikov, Y.V., Barshevskaya, T.N., Yavashev, L.P., Gurova, A.G. & Volkova, L.I. (1981) *Eur J Biochem* **117**: 195-200.

Athauda, S.B.P., Tanji, M., Kageyama, T. & Takahashi, K. (1989) A comparative study on the N-terminal amino acid sequences and some other properties of six isozymic forms of human pepsinogens and pepsins. *J Biochem* **106**: 920-927.

Auer, H.E. & Glick, D.M. (1984) Early events of pepsinogen activation. *Biochemistry* **23**: 2735-2739.

Baker, E.N. & Hubbard, R.E. (1984) Hydrogen bonding in globular proteins. *Prog Bioph Mol Biol* **44**: 97-179.

Bartlett, P.A. & Kazer, W.B. (1984) Phosphonic acid dipeptide analogues: potent, slow-binding inhibitors of aspartic peptidases. *J Am Chem Soc* **106**: 4282-4283.

Bartlett, P.A. & Marlowe, C.K. (1983) Phosphomamidates as transition-state analogue inhibitors of thermolysin. *Biochemistry* **22**: 4618-4624.

Bartlett, P.A., Hanson, J.E. & Giannousis, P.P. (1990) Potent inhibition of pepsin and penicillopepsin by phosphorus-containing peptide analogues. *J Org Chem* **55**: 6268-6274.

Bateman, K.S., Cherney, M.M., Tarasova, N.I., James, M.N.G. (1998) Crystal structure of human pepsinogen A. In *The Aspartic Proteinases: Retroviral,*

Fungal, Plant and Mammalian (ed. James, MNG; Plenum Press, NY), in the press.

Baudys, M. & Kostka, V. (1983) Covalent structure of chicken pepsinogen. *Eur J Bioch* **136**: 89-99.

Bernal, J.D. & Crowfoot, D. (1934) X-ray photographs of crystalline pepsin. *Nature* **133**: 794-795.

Cutfield, S.M., Dodson, E.J., Anderson, B.F., Moody, P.C., Marshall, C.J., Sullivan, P.A. & Cutfield J.F. (1995) The crystal structure of a major secreted aspartic proteinase from Candida albicans in complexes with two inhibitors. *Structure* **3**: 1261-1271.

Davie, E.W., Fujikawa K. & Kisiel, W. (1991) The coagulation cascade: initiation, maintenance and regulation. *Biochemistry* **30**: 10363-10370.

Davies, D.R. (1990) The structure and function of the aspartic proteinases. *Ann Rev Bioph Bioph Chem* **19**: 189-215.

Defize, J., Pals, G., Pronk, J.C., Frants, R.R., Rimmelzwaan, G., Westerveld, B.D. & Eriksson, A.W. (1985) Purification of pepsinogen A isozymogens by means of high resolution ion-exchange chromatography. Evidence for post-translational modifications. *Scand J Clin Lab Invest* **45**: 649-655.

Ding, J., Fraser, M.E., Meyer, J.H., Bartlett, P.A. & James, M.N.G. (1998) Macrocyclic inhibitors of penicillopepsin. 2. X-ray crystallographic analyses of penicillopepsin complexed with a P3-P1 macrocyclic peptidyl inhibitor and with its two acyclic analogues. *J Am Chem Soc* **120**: 4610-4621.

Dreyer, G.B., Metcalf, B.W., Tomaszek, T.A.J., Carr, T.J, Chandler, A.C.I, Hyland, L., Fakhoury, S.A., Magaard, V.W., Moore, M.L. & Strickler, J.E. (1989) Inhibition of human immunodeficiency virus 1 protease *in vitro*: rational design of substrate analogue inhibitors. *Proc Natl Acad Sci (USA)* **86**: 9752-9756.

Dunn, B.M. & Fink, A.L. (1984) Cryoenzymology of porcine pepsin. *Biochemistry* **23**: 5241-5247.

Dunn, B.M., Oda, K., Kay, J., Rao-Naik, C., Lowther, W.T., Beyer, B.M., Scarborough, P.E. & Bukhtiyarova, M. (1998) Comparison of the specificity of the aspartic proteinases towards internally consistent sets of oligopeptide substrates. *Adv Exp Med Biol* **436**: 133-138.

Dunn, B.M., Scarborough, P.E., Davenport, R. & Swietnicki, W. (1994) Analysis of proteinase specificity by studies of peptide substrates. *Methods Mol Biol* **36**: 225-243.

Edwards, D.R., and Murphy, G. (1998) Cancer. Proteases--invasion and more. *Nature* **394**: 527-528.

Esnouf, R.M. (1997) An extensively modified version of MolScript that includes greatly enhanced colouring capabilities. *J Mol Graphics* **15**: 133-138.

Foltmann, B. (1981) Gastric proteinases - structure, function, evolution, and mechanism of action. *Essays Biochem* **17**: 52-84.

Francis, S.E., Banerjee, R. & Goldberg, D.E. (1997) Biosynthesis and maturation of the malarial aspartic hemoglobinases plasmepsins I and II. *J Biol Chem* **272**: 14961-14968.

Fraser, M.E., Strynadka, N.C.J., Bartlett, P.A., Hanson, J.E. & James, M.N.G. (1992) Crystallographic analysis of transition-state mimics bound to penicillopepsin: Phosphorus-containing peptide analogues. *Biochemistry* **31**: 5201-5214.

Fruton, J.S. (1976) The mechanism of catalytic action of pepsin and related acid proteinases. *Adv Enzymol* **44**: 1-36.

Fujinaga, M., Chernaia, M.M., Tarasova, N.I., Mosimann, S.C. & James, M.N.G. Crystal structure of human pepsin and its complex with pepstatin. *Protein Sci.* **4**: 960-972 (1995).

Fusek, M. & Vetvicka, V. (1995) In *Aspartic Proteinases: Physiology and Pathology*. (CRC Press, Boca Raton, Florida).

Germaine, G.R. & Tellefson, L.M. (1981) Effect of pH and human saliva on protease production by *Candida albicans*. *Infect Immun* **31**: 323-326.

Greenberg, AH. (1996) Granzyme B-induced apoptosis. *Adv Exp Med & Biol 406*: 219-228.

Groettrup, M., Soza, A., Jucjelkorn, U. and Kloetzel, P.M. (1996) Peptide antigen production by the proteasome: complexity provides efficiency. *Immunology Today* **17**: 429-435.

Hayano, T., Sogawa, K., Ichihara, Y., Fuji-Kuriyama, Y. & Takahashi, K. (1988) Primary structure of human pepsinogen C gene. *J Biol Chem* **263**: 1382-1385.

Herriott, R.M. (1938) Isolation, crystallization and properties of swine pepsinogen. *J Gen Physiol* **21**: 501-540.

Herriott, R.M. (1939) Kinetics of the formation of pepsin from swine pepsinogen and identification of an intermediate compound. *J Gen Physiol* **22**: 65-78.

Hofmann, T. (1989) *Zoological Research*, Beijing (in Chinese).

Hofmann, T., Fink, A.L. & Dunn, B.M.. (1984) Cryoenzymology of penicillopepsin. Mechanism of action of aspartyl proteinases. *Biochemistry* **23**: 5247-5256.

Hofmann, T., Allen, B., Bendiner, M., Blum, M. & Cunningham, A. (1988) Effects of secondary substrate binding in penicillopepsin: contributions of subsites S3 and S2' to kcat. *Biochemistry* **27**: 1140-1146.

Hsu, I.-N., Delbaere, L.T.J. & James, M.N.G. (1977) Penicillopepsin from *Penicillium janthinellum* crystal structure at 2.8 Å and sequence homology with porcine pepsin. *Nature* **266**: 140-145.

Ichihara, Y., Sogawa, K., Morohashi, K., Fuji-Kuriyama, Y. & Takahashi, K. (1986) Nucleotide sequence of a nearly full-length cDNA coding for pepsinogen of rat gastric mucosa. *Eur J Biochem* **161**: 7-12.

International Union of Biochemistry and Molecular Biology. (1992) Enzyme nomenclature: recommendations of the nomenclature committee of the International Union. (Academic Press, San Diego).

James, M.N.G. & Sielecki, A.R. (1983) Structure and refinement of penicillopepsin at 1.8 Å resolution. *J Mol Biol* **163**: 299-361.

James, M.N.G. & Sielecki, A.R. (1985) X-ray diffraction studies on penicillopepsin and its complexes: the hydrolytic mechanism. In *Aspartic Proteinases and their Inhibitors* (V. Kostka, ed.; Walter de Gruyter, New York, 1985), pp. 163-177.

James, M.N.G. & Sielecki, A.R. (1986) Molecular structure of an aspartic proteinase zymogen, porcine pepsinogen, at 1.8Å resolution. *Nature (London)* **319**: 33-38.

James, M.N.G., Sielecki, A.R., Hayakawa, K. & Gelb, M.H. (1992) Crystallographic analysis of transition state mimics bound to penicillopepsin: difluorostatine- and difluorostatone-containing peptides. *Biochemistry* **31**: 3872-3886.

Kageyama, T. & Takahashi, K. (1986a) The complete amino acid sequence of monkey progastricsin. *J Biol Chem* **261**: 4406-4419.

Kageyama, T. & Takahashi, K. (1986b) The complete amino acid sequence of monkey pepsinogen A. *J Biol Chem* **261**: 4395-4405.

Kageyama, T., Ichinose, M., Miki, K., Athauda, S.B., Tanji, M. & Takahashi, K. (1989) Difference of activation processes and structure of activation peptides in human pepsinogens A and progastricsin. *J Biochem* **105**: 15-22.

Kageyama, T., Ichinose, M., Tsukada, S., Miki, K., Kurokawa, K., Koiwai, O., Tanji, M., Yakabe, E., Athauda, S.B.P. & Takahashi, K. (1992) Gastric procathepsin E and progastricsin from guinea pig. *J Biol Chem* **267**: 16450-16459.

Khan, A.R. & James, M.N.G. (1998) Molecular mechanisms for the conversion of zymogens to active proteolytic enzymes. *Protein Sci* **7**: 815-836.

Klionsky, D.J., Banta, L.M. & Emr S.D. (1988) Intracellular sorting and processing of a yeast vacuolar hydrolase: proteinase A propeptide contains vacuolar targetting information. *Mol Cell Biol* **8**: 2105-2116.

Kothary, M.H., Chase, T. Jr. & Macmillan, J.D. (1984) Correlation of elastase production by some strains of Aspergillus fumigatis with ability to cause pulmonary invasive aspergillosis in mice. *Infect Immun* **43**: 320-325.

Kwon-chung, K.J., Lehman, D., Good, C. & Magee, P.T. (1985) Genetic evidence for role of extracellular proteinase in virulence of *Candida albicans*. *Infect Immun* **49**: 571-575.

Kunitz, M. (1938). Formation of trypsin from trypsinogen by an enzyme produced by a mold of the genus *Penicillium*. *J Gen Physiol* **21**: 601-620.

Marciniszyn, J. Jr., Huang, J.S., Hartsuck, J.A. & Tang, J. (1976) Mechanism of intramolecular activation of pepsinogen. *J Biol Chem* **251**: 7095-7102.

McPhie, P. (1972) A spectrophotometric investigation of the pepsinogen-pepsin conversion. *J Biol Chem* **247**: 4277-4281.

Miller, M., Jaskolski, M., Rao, J.K.M., Leis, J. & Wlodawer, A. (1989) Crystal structure of a retroviral protease proves relationship to aspartic protease family. *Nature* **337**: 576-579.

Monod, M., Togni, G., Hube, B. & Sanglard, D. (1994) Multiplicity of genes encoding secreted aspartic proteinase in *Candida* species. *Mol Microbiol* **13**: 357-368.

Moore, S.A., Sielecki, A.R., Chernaia, M.M., Tarasova, N.I. & James, M.N.G. (1995) Crystal and molecular structures of human progastricsin at 1.62Å resolution. *J Mol Biol* **247**: 466-485.

Navia, M.A., Fitzgerald, P.M.D., Mckeever, B.M., Leu, C.-T., Heimbach, J.C., Herber, W.K., Sigal, I.S., Darke, P.L., Springer, J.P. (1989) Three-dimensional structure of aspartyl protease from human immunodeficiency virus HIV-1. *Nature* **337**: 615-620.

Nielsen, F.S. & Foltmann, B. (1993) Activation of porcine pepsinogen A: the stability of two non-covalent intermediates at pH 8.5. *Eur J Biochem* **217**: 137-142.

Olsen, V., Guruprasad, K., Cawley, N.X., Chen, H.-C., Blundell, T.L. & Loh, Y.P. (1998) Cleavage efficiency of the novel aspartic protease Yapsin 3 (Yap3p) enhanced for substrates with arginine residues flanking the P1 site: correlation with electronegative active-site pockets predicted by molecular modeling. *Biochemistry* **37**: 2768-2777.

Pearl, L.H. (1987) The catalytic mechanism of aspartic proteinases. *FEBS Lett* **214**: 8-12.

Rao, C. & Dunn, B.M. (1995) Evidence for electrostatic interactions in the S2 subsite of porcine pepsin. *Adv Exp Med Biol* **362**: 91-94.

Samloff, I.M. (1969) Slow moving protease and the seven pepsinogens. *Gastroenterology* **57**: 659-669.

Samloff, I.M. & Liebmann, W.M. (1973) Cellular localization of the Group II pepsinogens in human stomach and duodenum by immunofluorescence. *Gastroenterology* **65**: 36-42.

Schechter, I., Berger, A. (1967) On the size of the active site in proteases. I. Papain. *Biochem Biophys Res Commun* **27**: 157-162.

Sielecki, A.R., Hayakawa, K., Fujinaga, M., Murphy, M.E., Fraser, M., Muir, A.K., Carilli, C.T., Lewicki, J.A., Baxter, J.D., James, M.N.G. (1989) Structure of recombinant human renin, a target for cardiovascular-active drugs, at 2.5Å resolution. *Science* **243**: 1346-1351.

Sodek, J. & Hofmann, T. (1970) Amino acid sequence around the active site aspartic acid in penicillopepsin. *Can J Biochem* **48**: 1014-1016.

Sogawa, K., Fuji-Kuriyama, Y., Mizukami, Y., Ichihara, Y. & Takahashi, K. (1983) Primary structure of human pepsinogen gene. *J Biol Chem* **258**: 5306-5311.

Suguna, K., Padlan, E.A., Smith, C.W., Carlson, W.D. & Davies, D.R. (1987) Binding of a reduced peptide inhibitor to the aspartic proteinase from *Rhizopus chinensis*: implications for mechanism of action. *Proc Natl Acad Sci (USA)* **84**: 7009-7013.

Taggart, R.T., Mohandas, T.K., Shows, T.B. & Bell, G.I. (1985) Variable numbers of pepsinogen genes are located in the centromeric region of human chromosome 11 and determine the high frequency electrophoretic polymorphism. *Proc Natl Acad Sci (USA)* **82**: 6240-6244.

Tang J, James MNG, Hsu IN, Jenkins JA, Blundell TL. 1978. Structural evidence for gene duplication in the evolution of the acid proteases. *Nature* **271**: 618-621.

Thangamani, A. & Hofmann, T. (1966) The role of a protease in sporulation of *Penicillium janthinellum*. *Can J Biochem* **44**: 579-584.

Togni, G., Sanglard, D. & Monod, M. (1994) Acid proteinase secreted by *Candida tropicalis*: virulence in mice of proteinase negative mutant. *J Med Vet Mycol* **32**: 257-265.

Tronrud, D.E., Holden, H.M. & Matthews, B.W. (1987) Structures of two thermolysin-inhibitor complexes that differ by a single hydrogen bond. *Science* **235**: 571-574.

Twining, S.S., Sealy, R.C. & Glick, D.M. (1981) Preparation and activation of spin-labeled pepsinogen. *Biochemistry* **20**: 1267-1272.

Vallet, V., Ahmed, C., Gaeggler H.-P., Horisberger, J.-D. and Rossier, B.C. (1997) An epithelial serine protease activates the amiloride-sensitive sodium channel. *Nature* **389**: 607-610.

Weaver, L.H., Kester, W.R. & Matthews, B.W. (1977) A crystallographic study of the complex of phosphoramidon with thermolysin. *J Mol Biol* **114**: 119-132.

Weinmaster, G. (1998) Reprolysins and astacins...alive, alive-o. *Science* **279**: 336-337.

Westphal, V., Marcusson, E.G., Winther, J.R., Emr, S.D. & van den Hazel, H.B. 1996. Multiple pathways for vacuolar sorting of yeast proteinase A. *J Biol Chem* **271**: 11865-11870.

Wlodawer, A. & Erickson, J.W. (1993) Structure-based inhibitors of HIV-1 protease. *Ann Rev Biochem* **62**: 543-585.

CHAPTER 2

STRUCTURES OF CYCLIC AND ACYCLIC PEPTIDE-BASED PHOSPHONATE INHIBITORS WITH PENICILLOPEPSIN

A. INTRODUCTION

The synthesis of potent and specific inhibitors of biologically important proteolytic enzymes is a widespread approach for rational drug design. In this strategy, a "lead" compound that inhibits an enzyme of interest is typically identified by a combinatorial search of small molecules, or by past experience with related enzymes and their inhibitors (Salemme et al., 1997). Subsequently, an important step is the improvement of the inhibitor against the proteolytic enzyme by inspection of the X-ray structure of the enzyme-inhibitor complex. The inhibitor can be rendered more potent by optimizing the hydrophobic, hydrogen bonding, and electrostatic interactions with the enzyme active site.

A further enhancement of the binding constant can be achieved by stabilizing the structure that the inhibitor adopts upon binding the active site, assuming that the inhibitor is more flexible in aqueous solution. As previously mentioned, peptide-based inhibitors of aspartic proteinases typically interact with the active site cleft in an extended β-strand conformation. The entropic cost associated with the loss of conformational flexibility for these inhibitors would be minimized upon stabilization of this conformation in solution. These energetic considerations have led to the development of phosphonate-based inhibitors of penicillopepsin that are covalently linked at their side chains in order to constrain their conformation and enhance the binding affinity (Meyer & Bartlett, 1998; Smith & Bartlett, 1998). The design of these inhibitors was guided by the structure of penicillopepsin in complex with an

acyclic phosphonate-based peptide analogue, hereafter referred to as 'PP2' (penicillopepsin phosphonate inhibitor 2; Fraser et al., 1992; Fig. 2A.1).

Structure of the complex PP2-penicillopepsin. Previous crystallographic analysis of the phosphorus-containing peptide-mimic PP2 revealed that the inhibitor binds in an extended fashion across the catalytic cleft of penicillopepsin (Fig. 2A.2). The phosphonate moiety of PP2 is co-ordinated between the two catalytic Asp residues, Asp33 and Asp213, and is unable to be hydrolyzed by the enzyme. The measured inhibition constant, K_i, for the complex of PP2 with penicillopepsin is 2.8 nM (Bartlett et al., 1990). The tetrahedral geometry at the phosphorus atom resembles the cleavage transition state that would be expected for a genuine substrate. The *pro-S* oxygen atom (O4) makes a 2.4 Å hydrogen bond to Asp33(Oδ1), shorter than the minimum 2.5 Å (and average 2.8 Å) hydrogen bonded contacts that are typical of proteins (Baker & Hubbard, 1984). The short contact distance and the geometry of the interaction has been interpreted as a hydrogen bond with a symmetric single-well potential energy function (Fraser et al., 1992), similar to very short hydrogen bonds that have been reported in the structures of thermolysin and carboxypeptidase in complex with phosphonate inhibitors (Holden et al., 1987; Kim & Lipscomb, 1991). The *pro-R* oxygen atom of PP2 is the recipient of a hydrogen bond (2.6 Å) from Asp213(Oδ1) of penicillopepsin, suggesting that this side chain is fully protonated (Fraser et al., 1992).

PP2 is involved in extensive hydrophobic and hydrogen bonding contacts with the substrate-binding cleft. Non-polar contacts include interactions between the side chains of P4, P2, P1 and P1' and the corresponding subsites S4, S2, S1 and S1' of penicillopepsin. In addition, main-chain hydrogen bonds between PP2 and penicillopepsin [P1(NH) and Gly215(CO), P1'(CO) and Gly76(NH)] stabilize the complex further, and are indicative of the intimate contacts that are characteristic of

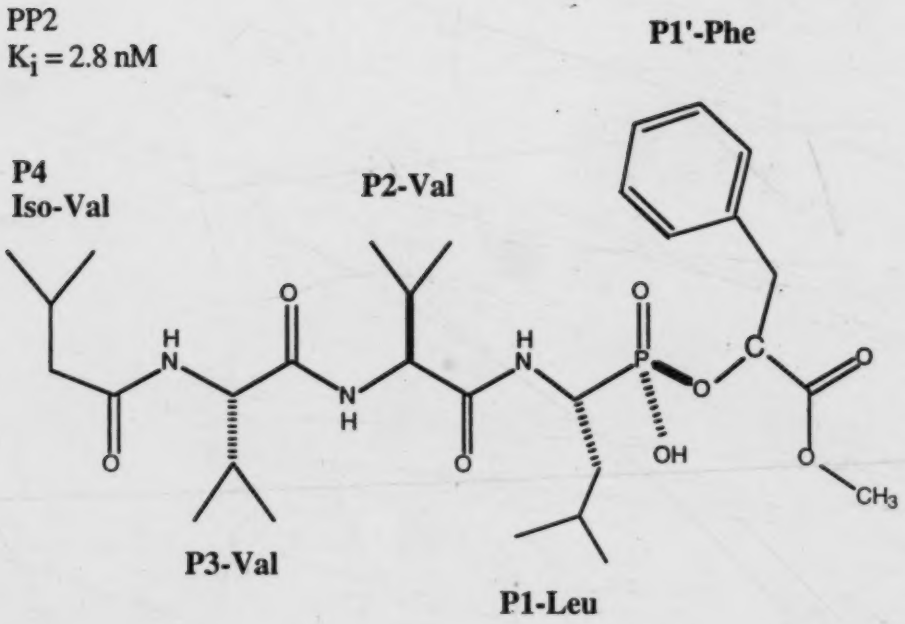


Figure 2A.1 : Chemical drawing of PP2, the phosphonate-based peptide inhibitor of penicillopepsin.

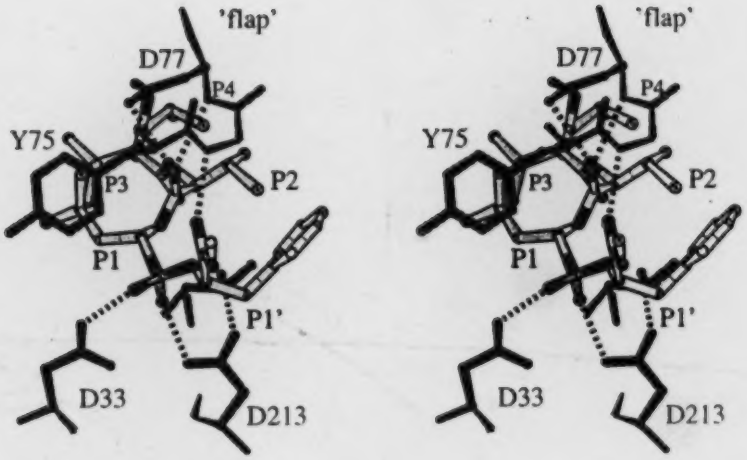
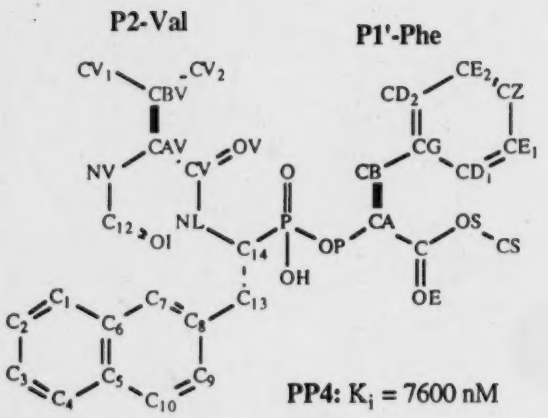


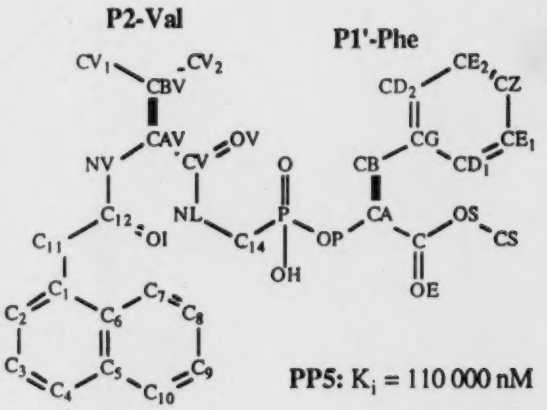
Figure 2A.2: Active site structure of the PP2-penicillopepsin complex (Fraser et al., 1992; PDB code 1ppl). The inhibitor is distinguishable from the enzyme by the lighter shading of bond vectors. The side chains of PP2 are labelled according to the nomenclature of Shechter and Berger (1967). Hydrogen bonds are shown with dashed lines. The side chain Og atom of Thr216 (unlabelled) forms a hydrogen bond with Asp213 at the active site. The figures in this chapter were produced by BOBSCRIPT (Esnouf, 1997).

aspartic proteinases with their inhibitors (reviewed by Davies, 1990). As a consequence of the β-strand conformation of PP2, it was noticed that alternating pairs of side chains (P1---P3, P2---P1') are relatively close together in space. These two pairs of side chains interact with hydrophobic parts of the substrate-binding cleft at opposite sides of the peptide, and there is sufficient space to connect each pair together. This strategy was first suggested by Blundell and co-workers to improve the potency and selectivity of renin inhibitors (Blundell et al., 1986), although the proposal was never followed through experimentally. These considerations have led to the development of 'second-generation' phosphonate inhibitors with covalent bridges at alternating side chains to reduce their conformational freedom and thereby enhance the binding affinity. The macrocyclic inhibitors, along with their acyclic analogs (Fig. 2A.3), were designed to differentiate the entropic cost of binding a peptide to the active site from the favorable hydrophobic and polar interactions. In order to derive a valid quantitative estimate, it was necessary to demonstrate that the cyclic and acyclic inhibitors bind penicillopepsin in an identical manner.

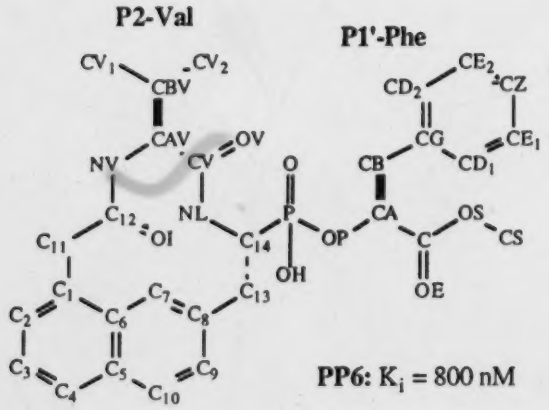
Structures of cyclic and acyclic phosphonate inhibitors with penicillopepsin. The first series of inhibitors were designed to assess the effects of a covalent link between the P3 and P1 side chains (Meyer & Bartlett, 1998). The chemical structure of the P3-P1 macrocyclic inhibitor (PP6) and its acyclic control compounds (PP4, PP5) are shown in Fig. 2A.3, together with the corresponding K_i values. The program "CAVEAT" was implemented to identify structural frameworks that would ideally link the side chains, followed by modeling and energy minimization procedures to evaluate the resulting possibilities (Lauri & Bartlett, 1994; Meyer & Bartlett, 1998). The naphthalene bridge shown in Fig. 2A.3 (PP4) was selected for its synthetic accessibility and rigidity, while the corresponding acyclic derivatives PP5 and PP6 lacked the methylene link at P3 and P1, respectively. Kinetic analyses revealed that



PP4: K_i = 7600 nM



PP5: K_i = 110 000 nM



PP6: K_i = 800 nM

Figure 2A.3: Chemical structures of 'second-generation' phosphonate inhibitors of penicillopepsin. These series of inhibitors have been designed to study the effects of linking the P3-P1 positions upon the binding affinity. A naphthyl bridge links the P3 and P1 side chains in PP6.

the most conformationally constrained compound, PP6, was the most potent inhibitor (K_i = 800 nM). The K_i values for the control acyclic analogues, PP4 (7.6 μM) and PP5 (110 μM), were significantly higher, suggesting that binding affinity increases with decreasing conformational freedom (Meyer & Bartlett, 1997).

The inhibitors were synthesized in sufficient quantities to enable co-crystallization with penicillopepsin for structural analyses (Ding et al., 1998). Table 2A.1 gives a brief account of the resolution and quality of the various structures. The structure of PP6-penicillopepsin revealed that the macrocyclic inhibitor interacts with the enzyme in an almost identical manner relative to PP2, the acyclic parent compound (Fig. 2A.4). The phosphonyl oxygen atoms of PP6 form hydrogen bonds to Asp33 and Asp213, and the flap is in a 'closed' position such that P2-Val(O) receives hydrogen bonds from Asp77(NH) and Gly76(NH). Despite these similarities, the naphthalene ring that links the P1 and P3 side chains was located in a more polar environment relative to PP2, making several contacts (< 3.6 Å) with water molecules and polar residues within the flap (Tyr75, Asp77, Ser79).

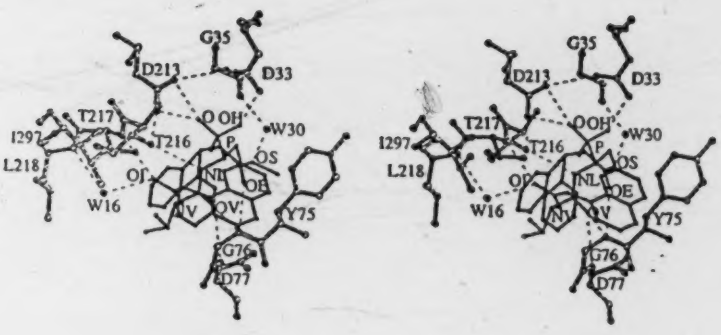
Unfortunately for the proof of the concept, the structure of the PP4-penicillopepsin complex, in which the P1(C_α) atom of the acyclic inhibitor branches into a naphthylmethyl substituent, revealed significant structural differences when compared to PP6 (Ding et al., 1998). The bound naphthalene ring of PP4 was rotated at the χ_2 angle by -60° relative to PP6, thereby adopting a χ_2 angle of 30° (Fig. 2A.4). This side chain conformation is relatively unfavourable when compared to χ_2 angles that are typically observed for aromatic residues (Phe, Tyr, Trp) in the protein database (Schrauber et al., 1993). However, adoption of the unfavorable χ_2 value was compensated by additional hydrophobic interactions between the naphthyl ring and Phe112, thus bringing the P1 residue into more intimate contacts with the S1 binding pocket (Ding et al., 1998). Omitting the naphthalene and phenyl rings, a least-squares superposition of the equivalent backbone atoms of PP4 and PP6 yielded

Table 2A.1: Structures of Cyclic and Acyclic Phosphonate Inhibitors in Complex with Penicillopepsin

Inhibitor	PDB code	Resolution (Å)	Temp (°K)	†R-factor (R-free)
PP2	1ppl	1.70	295	13.1%
PP4	2wea	1.25	110	15.5% (19.2)
PP5	2web	1.50	295	16.2% (19.0)
PP6	2wec	1.50	295	16.5% (21.2)

†$R=\Sigma||F_o|-|F_c||/\Sigma|F_o| \times 100\%$

(a)



(b)

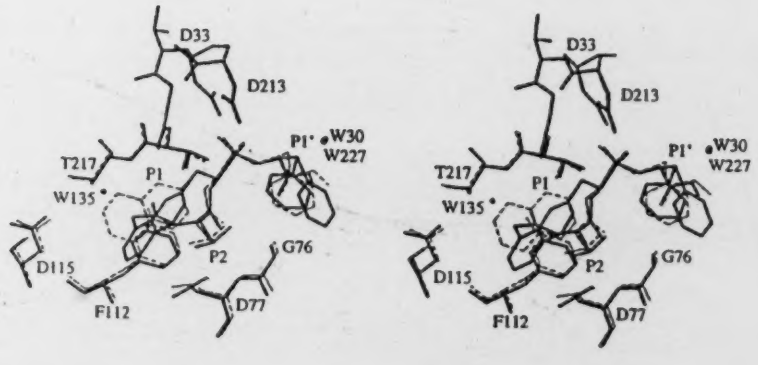


Figure 2A.4: Structures of the P3-P1 macrocyclic inhibitor and its acyclic control with penicillopepsin. (a) Stereoview of the PP6-penicillopepsin complex at the active site (from Ding et al., 1998). The enzyme is lightly shaded, relative to the inhibitor. Hydrogen bonds are represented by dashed lines. The P1'-Phe ring is shown as two alternate conformers. (b) Superposition of PP6-penicillopepsin (solid lines) with acyclic PP4-penicillopepsin (dashed lines). A rotation of the naphthyl ring of PP4 is apparent, placing it in two very different environments. However, the architecture of the enzyme active site is unchanged.

a relatively low rms displacement of 0.309 Å for 22 atoms. Similarly, a least-squares superposition of penicillopepsin in the two complexes resulted in an rms difference of only 0.205 Å for 1248 backbone atoms (Ding et al., 1998). Comparisons of the substrate-binding clefts revealed negligible changes in the enzyme structure between the two complexes, so that the major difference between the two complexes was the conformation of the side chain naphthyl moiety of the inhibitors. Overall, these observations suggest that the enzyme provides a rigid framework at the active site, but the inherent flexibility of the inhibitors leads to significant differences in their bound structures as multiple conformations are sampled during formation of the complex.

Attention was switched to the remaining acyclic control, PP5, in which the naphthalene ring is linked to the P3 residue, leaving a glycine at the P1 position. Disappointingly, the structure of PP5-penicillopepsin revealed even more profound differences in relation to PP6-penicillopepsin (not shown). Superposition of the bound structures of PP5 and PP6, following alignment of penicillopepsin, yielded an rms deviation of 2.08 Å for 32 common atoms (Ding et al., 1998). Surprisingly, analysis of the PP5-penicillopepsin structure showed that the phosphonyl group was shifted about 3 Å away from the two catalytic aspartate residues, Asp33 and Asn213 (Ding et al., 1998). Instead, a water molecule was observed between the two catalytic Asp residues, in the equivalent position that is assumed by the catalytic water molecule in native penicillopepsin (James et al., 1983). The naphthyl moiety of PP5, which should reside in the S3 pocket of pencillopepsin, had moved into the S4 pocket. These overall structural changes resulted in fewer hydrogen bonding and hydrophobic contacts, and likely contributed to the poor inhibition constant for PP5 (K_i = 110 μM). It was speculated that perhaps the complex was an "intermediate" along the binding pathway such that the presence of the large naphthyl group prevented the peptide from sliding into the thermodynamically favored position that is usually observed for phosphonate inhibitors (Ding et al., 1998). While the unique

position of PP6 at the active site was intriguing, this acyclic inhibitor failed to provide an adequate control for the PP4-penicillopepsin complex.

In summary, the first series of cyclic and acyclic inhibitors designed around linking the P3 and P1 side chains were sufficiently variable in their interactions with penicillopepsin to be rendered inappropriate for quantitation of the entropic cost of peptide binding. Consequently, a second macrocyclic inhibitor was designed and synthesized (Smith & Bartlett, 1998) to exploit the spatial proximity of the P2 and P1' side chains (Fig. 2A.5). The complex PP7-penicillopepsin and its acyclic control PP8-penicillopepsin have been crystallized and their structures refined. The striking similarities of these two structures, and with their parent complex PP2-penicillopepsin, demonstrate that these inhibitors are suitable for quantifying the effects of conformational constraints on the binding affinity.

PP7: X = CH_2 K_i = 0.10 nM

PP8: X = H,H K_i = 41.8 nM

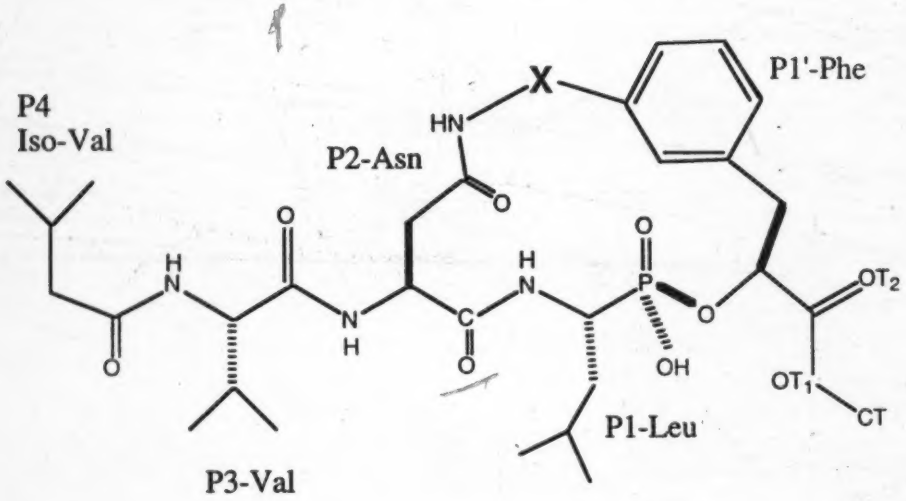


Figure 2A.5: Chemical drawing of P2-P1' macrocycle and its acyclic control.

B. MATERIALS AND METHODS

Crystallization of PP7-inhibitor complex. The macrocyclic peptide-based inhibitor PP7 {Methyl Cyclo[(2S)-2-[[(1R)-1-(N-L-N-(3-methylbutanoyl)valyl-L-aspartyl) amino)-3-methylbutyl]-hydroxyphosphinyloxy]-3-(3-aminomethyl)phenyl-propanoate] lithium salt} and its acyclic derivative PP8 were synthesized as published previously (Smith & Bartlett, 1998). The inhibitors were obtained by our laboratory as lyophilized powders. Penicillopepsin was obtained as a powder preparation from Professor T. Hofmann, University of Toronto (Fraser et al., 1992).

The powder samples were solubilized in separate aqueous solutions containing 20 mM sodium acetate (pH 4.6). The solutions were then mixed to form a stock solution of the enzyme-inhibitor complex, with a penicillopepsin:PP7 molar ratio of 1:6. Crystallization of the complex was performed using the hanging drop technique following mixture of equivalent (~5 μL) volumes of the enzyme-inhibitor solution with a solution containing ammonium sulfate as the precipitant. The final concentrations in the drop were about 12 mg/mL of protein, 35% saturated ammonium sulfate, and 100 mM sodium acetate (pH 4.6). Micro-crystals appeared within 24 hours, and these were used as seeds to obtain larger crystals. Typically, one further round of macro-seeding was employed to increase crystal size, producing monoclinic plate-like crystals (0.15 x 0.5 x 0.5 mm) within 2 weeks.

A similar procedure to that described above was followed to grow crystals of the PP8-penicillopepsin complex (Dr. Jonathan Parrish, unpublished).

Data collection and processing. Low temperature data collection involved transfer of the crystals to a cryoprotectant containing 25% glycerol. The procedure involved a careful mixing of the drop containing the crystal with an equal amount of the precipitant solution that was supplemented with glycerol. The transfer was performed gradually over 30-40 minutes, in 5% v/v increments of glycerol up to 25% v/v, with

several transfers at each concentration. This protocol was found to be ideal for preventing the crystal from cracking and thereby maintaining the excellent quality of diffraction. Crystals were mounted onto nylon loops (Cryoloops, Hampton Research) and placed in a stream of nitrogen gas maintained at 105°K.

In the laboratory, X-rays were generated using a Siemens rotating anode X-ray generator operating at 40 kV and 150 mA, and reflections were recorded with a Siemens X1000 multiwire area detector system. Data to 1.43 Å resolution were processed with XENGEN (Howard et al., 1987), and scaled using SCALA (CCP4 package; Evans, 1993). Synchrotron data to a maximum of 0.90 Å resolution were collected at the Cornell High Energy Synchrotron Source (CHESS, Ithaca, New York), processed using MOSFLM (Leslie, University of Cambridge), and scaled using SCALA (Evans, 1993). In the case of the initial refinement using X-PLOR, structure factor amplitudes and standard deviations were obtained using TRUNCATE (French & Wilson, 1978), as implemented in the CCP4 package. Details of the crystal parameters and quality of the data are provided in Table 2B.1.

Refinement of the models. The structure of the PP7-penicillopepsin complex was initially refined using the model of PP2-penicillopepsin [Fraser et al., 1992; Protein DataBank (PDB) code 1ppl]. The starting model was retrieved from the database and all waters, sugars and other small molecules were deleted from the atom coordinate list. Prior to the first step, 10% of the reflections were randomly assigned an 'R-free' flag and excluded from the refinement process. The polypeptide (residues 1-323) was subjected to rigid body minimization and simulated annealing using a least-squares target function implemented in X-PLOR (Brunger, 1993), using data between 8-2.5 Å resolution. At this stage, a SIGMAA-weighted electron density map (Read, 1986) was calculated using the program from the CCP4 package. The $2F_o$-F_c and F_o-F_c maps revealed clear density for all of the atoms corresponding to the inhibitor

Table 2B.1: Crystal Data for PP7 and PP8 Complexes with Penicillopepsin

	PP7-penicillopepsin		PP8-penicillopepsin
	(a) CHESS, 100°K	(b) Lab, 295°K	Lab, 105°K
Space Group			
a (Å)	96.98	97.86	96.24
b (Å)	46.65	46.57	46.47
c (Å)	65.71	66.57	65.38
β (°)	115.57	115.93	115.58
X-rays	0.919 Å	1.5418 Å	1.5418 Å
detector	2 x 2 CCD	DIP image plate detector	Siemens multiwire area detector
max resolution on detector	0.90 Å	1.55 Å	1.45 Å
number of measurements	430728	115357	107304
number of unique reflections	160142	39959	45997
$^{\dagger}R_{merge}$	5.3%	6.5%	7.4%
Completeness	98.0% (20-0.95 Å)	97.3% (20-1.55 Å)	94.8% (8-1.44 Å)
	51.7% (0.97-0.95 Å)	74.0% (1.58-1.55 Å)	66.3% (1.46-1.44 Å)
$\langle F \rangle / \langle \sigma F \rangle$ greater than 3.0	39.1% (0.97-0.95 Å)	66.5% (1.58-1.55 Å)	39.1% (1.46-1.44 Å)

$^{\dagger}R_{merge} = \Sigma_{hklj} | I_{hklj} - \langle I_{hkl} \rangle | / \Sigma_{hklj} \langle I_{hkl} \rangle \times 100\%$

PP7, which was therefore included in the model. A set of distance and angle restraints derived from the topology and parameter files of standard amino acids (Engh & Huber, 1991), and the geometry corresponding to the previous structure of PP2-penicillopepsin (Fraser et al., 1992), were used to restrain the model of the PP7 macrocycle. Subsequently, several rounds of map fitting, restrained energy minimization (X-PLOR), and *B*-factor refinement were performed in concert with a gradual increase in resolution to include data between 10-1.1 Å. A bulk solvent correction in X-PLOR was applied to approximate the scattering by diffuse waters and its contribution to the low angle reflections. During these iterative cycles, water molecules, sulfate ions, and glyerol molecules were added to the model using the graphics program O (Jones & Kjeldgaard, 1995).

Upon reaching a reasonable crystallographic residual using X-PLOR (R=18.6%, R-free=19.6%), the program SHELXL (Sheldrick, 1997) was used for further refinement including all data between 10-1.1 Å (156181 reflections). Initially, two rounds of restrained conjugate gradient least-squares (CGLS) minimization were performed, each consisting of 10 cycles of positional and isotropic *B*-factor refinement. The minimization was performed against diffraction intensities that were obtained from SCALA (Evans, 1993). Water molecules that contained excessively high *B*-factors were removed. Folowing these two rounds, the conventional R-factor and R-free, corresponding to the structure factor amplitudes, were 17.86% (140539 data) and 19.2% (15642 data). Further rounds of refinement in SHELXL consisted of 10 cycles of positional and anisotropic *B*-factor refinement. After each round, 2Fo-Fc and Fo-Fc maps were calculated and the model was examined using XtalView (McCree, 1992). Alternate side chain conformations, additional waters, sulfate ions, and glycerol molecules were gradually included in the model. The various steps in the refinement are summarized in Table 2B.2.

Table 2B.2: Stages in the Refinement of PP7-Penicillopepsin

Stage	Resolution	reflections	H_2O	R_{cryst}	R_{free}	Comments
1	8-1.50 Å	42 235	----	39.9%	39.6%	-start refinement from PDB file 1ppl, PP2-penicillopepsin, stripped of waters, sugars, and the inhibitor -XPLOR simulated annealing, individual B-factors
2	8-1.50 Å	42 235	----	25.2%	27.2%	-add waters, positional refinement in XPLOR, extend resolution
3	8-1.30 Å	64 812	230	19.3%	20.9%	-add sulfates, glycerol, multiple conformations, sugars, extend resol.
4	10-1.10 Å	104 253	381	18.4%	19.6%	-move to SHELX-L, perform several rounds (10 cycles/round) of conjugate gradient L.S., map fitting, and extend resolution -remove waters, re-calculate peaks
5	10-0.90 Å	156 181	291	17.4%	19.2%	-introduce anisotropic B-factors
6	10-0.90 Å	156 181	370	12.34%	14.46%	-several rounds of CGLS -omit maps for disordered regions (loop 277-281, Thr-linked sugar)
7	10-0.90 Å	156 181	491	9.91%	12.47%	-2 rounds of CGLS (all data)
8	10-0.90 Å	156 181	491	9.85%	---	-one cycle of full-matrix least-squares

Refinement of the structure of PP8-penicillopepsin was initiated using a partially-refined model of PP7-penicillopepsin. The starting model was stripped of the waters, sulfates and glycerol prior to a rigid body minimization (X-PLOR, 8-2.5 Å data) that treated PP7 and penicillopepsin as separate molecules. Upon inspection of the 2Fo-Fc and Fo-Fc maps, the PP8 inhibitor was observed to be in a similar position at the active site cleft, relative to PP7. However, a break in continuity between the P2 (Asn) and P1' (Phe) side chains was evident, as expected for this acyclic analog. In addition, a rotation of the P2-Phe side chain, corresponding to about 30° in χ_2, was also observed. Map fitting (using the program O, Jones & Kjeldgaard, 1995; XtalView, McCree, 1992) and further refinement was performed as described for PP7-penicillopepsin, using all data between 8-1.43 Å. In the latter stages of refinement, several rounds of CGLS minimization (SHELXL) were performed using all data.

Finally, water molecules in the models of PP7-penicillopepsin and PP8-penicillopepsin were chosen from peaks greater than $4\sigma(I)$ and their proximity to hydrogen-bonding partners. Isotropic *B*-factors were refined for all waters with their occupancies fixed at 1.0. However, several waters in PP7-penicillopepsin showed "split peaks" that were refined as alternate positions having partial occupancies in the model.

C. RESULTS

Quality of the models. The structures of the enzyme-inhibitor complexes have been refined to low crystallographic residuals and good geometry and stereochemistry. Details of the final results in the refinement of PP7 and PP8 in complex with penicillopepsin are described in Table 2C.1. The average value of the isotropic *B*-factors is 6.67 $Å^2$ for the acyclic inhibitor, PP8 (all atoms). In comparison, the averaged *B*-value for the backbone atoms of the enzyme is 10.5 $Å^2$, showing that PP8 is tightly associated with the active site. Similarly, the isotropic-equivalent *B*-factors corresponding to the inhibitor PP7 are small, relative to penicillopepsin in the complex. The co-ordinate error was estimated to be less than 0.13 Å for the PP8-penicillopepsin structure using a SIGMAA plot (Read, 1986) and from the Luzzati method (Luzzati, 1952). For the high-resolution structure of PP7-penicillopepsin, a restrained full-matrix least-squares cycle was initially performed on a DEC Alpha clone running a 600 Mhz Alpha 21164 chip in order to obtain accurate coordinate errors from the inverse matrix. However, due to the excessive memory requirement for ~3000 atoms, the single cycle continued for over one month without finishing, at which point the calculation was terminated manually. To reduce memory requirments, the full-matrix cycle was performed by dividing the coordinates into four overlapping and equal blocks. The block matrix cycle is currently being performed and should give results that approximate the full-matrix refinement (George Sheldrick, personal communication).

Structure of PP8-penicillopepsin. The complex of the acyclic inhibitor with the enzyme reveals that the phosphonate moiety of PP8 resides between the active site residues, Asp33 and Asp213 (Fig. 2C.1,2a). The phosponate oxygens make several hydrogen bonds and contacts with the Asp residues, and the tetrahedral geometry of the phosphorus atom mimics the cleavage state of substrates at the scissile carbonyl-

Table 2C.1:Refinement Statistics for the Penicillopepsin-Phosphonate Inhibitor Complexes

parameters	PP7	PP8
resolution range (Å)	10.0 --- 0.90	8.0 --- 1.45
reflections (no cutoff)	156181	30576
R_{cryst}¥	9.91% (140539)	16.1% (27728)
R_{free} (10%)	12.47% (15642)	19.9% (2848)
no. of parameters	27767	11596
model:		
no. of protein residues	1-323	1-323
no. of inhibitor atoms	44	43
no. of waters	491	412
no. of sulfate anions	2	2
no. of glycerols	2	2
rms deviations (SHELX-97)		
bond lengths (Å)	0.017	0.010
1,3 distances	0.034	0.029
rms co-ordinate error	<0.05†	0.13§
average isotropic B-factors (Å^2):		
backbone	18.5	10.5
inhibitor (all atoms)	14.8	6.67
waters	28.1	23.4
Ramachandran plot		
most favorable regions	91.7%	93.0%
additional regions	8.3%	7.0%

¥$R=\Sigma\,||F_o|-|F_c||/\Sigma|F_o| \times 100\%$

§Estimated using SIGMAA (Read, 1986)

†Estimated from a Luzzati plot (Luzzati, 1952). A full-matrix cycle is currently being performed to obtain the estimated standard deviations (esds) of the coordinates.

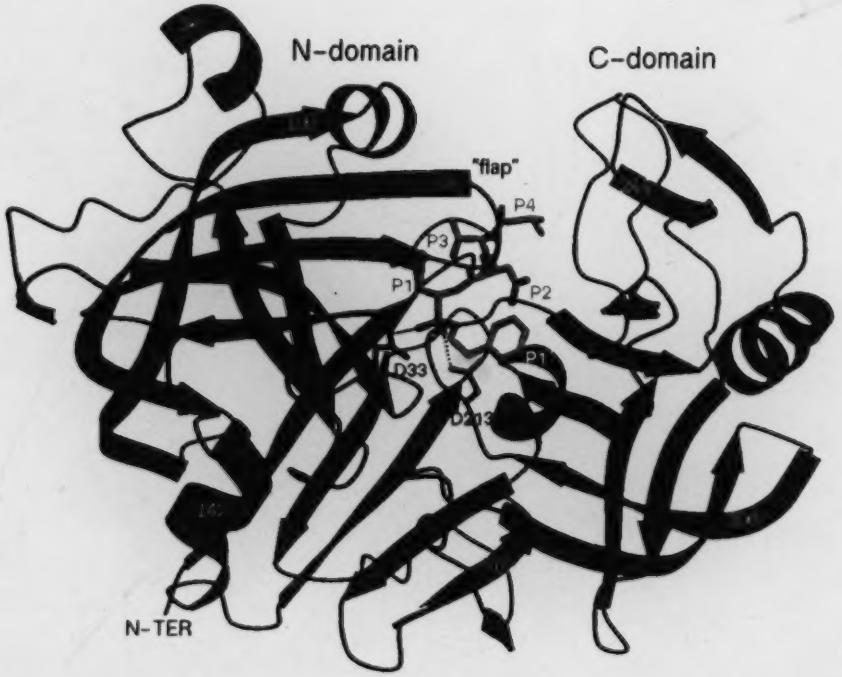


Figure 2C.1: Structure of the PP8-penicillopepsin complex. The view of the aspartic proteinase has been rotated 90° anti-clockwise along the line of view, relative to the introductory chapter. The orientation shown here is consistent with published structures of enzyme-inhibitor complexes. The P2 and P1' side chains of the inhibitor are close in space when the inhibitor is bound to the enzyme.

carbon. The *pro-S* oxygen atom (O4) of the phosphonate group makes a very short hydrogen bond (2.40 Å) to Asp33(Oδ2). The length and stereochemistry of this active site interaction is consistent with a hydrogen bond having a symmetric single-well potential energy function (Fraser et al., 1992; Ding et al., 1998). Table 2C.2 is a list of the hydrogen bonds between the enzyme and the inhibitors, and Table 2C.3 is a detailed analysis of the surface area that is buried upon association of the inhibitors with the enzyme. Table 2C.4 lists the conformational torsion angles of the various inhibitors in their complexes with penicillopepsin. Finally, Table 2C.5 provides the results of least-squares superpositions between the enzyme-inhibitor complexes in a matrix format. For completeness, the parent compound PP2 is also included in the analyses. As will be discussed shortly, the hydrogen-bonding and hydrophobic interactions, as well as the conformations of the three inhibitors bound at the active site of penicillopepsin are remarkably similar.

The side chains of PP8 (P4-P1') are observed in their corresponding enzyme subsites (S4-S1'), and both the side chains and the backbone of the inhibitor are involved in hydrogen-binding interactions with the enzyme (Fig. 2C.2). The "flap" region of penicillopepsin (residues 72-78) is observed in the "down" position, thus clamping the substrate to the active site. For example, Gly76(NH) of the flap makes a hydrogen bond to the carbonyl carbon of the P1' residue, which is a common interaction among complexes of aspartic proteinases with inhibitors (Ding et al., 1998; James et al., 1992; Fraser et al., 1992; Suguna et al., 1987; Davies, 1990). In addition, the P1-Leu side chain makes hydrophobic contacts with the hydrophobic S1 pocket, part of which is formed by the flap residue Tyr75. Overall, greater than 80% of the solvent-accessible surface area of PP8 is buried upon association with penicillopepsin. Interestingly, the side chain of P2-Asn(Nδ2) makes a 2.82 Å hydrogen bond to Thr216(Oγ1), which in turn is involved in a hydrogen bond *via* the Oγ1 atom with Asp213(Oδ2). This network of hydrogen bonds is indicative of the

intimate contacts between the inhibitor and the enzyme active site. A water molecule (W501) also participates in the association of PP8 with the enzyme through its hydrogen bonds with the carbonyl oxygen of P3-Val (2.89 Å) and with the backbone nitrogen of Asn218 (3.02 Å).

In summary, the structure of PP8-penicillopepsin reveals that the position and conformation of the inhibitor is consistent with its parent compound, PP2-penicillopepsin (Fraser et al., 1992; discussed in detail below). The tight binding of PP8 is a consequence of the extensive hydrogen bonding and hydrophobic interactions, as well as the transition-state mimicry of the non-cleavable phosphonate moiety at the active site. Thus, the structures of the majority of these second-generation phosphonate inhibitors in complexes with penicillopepsin have achieved these aspects of their design. The only exception is the complex of PP5-penicillopepsin (containing Gly at the P1 position), in which the phosphonate group has shifted away from the catalytic Asp residues and therefore the bound inhibitor does not resemble the binding of a transition-state analog.

Structure of PP7-penicillopepsin. The macrocyclic inhibitor PP7 has one of the lowest K_i values (0.10 nM) against an aspartic proteinase that has been reported in the literature. The complex PP7-penicillopepsin was refined to 0.95 Å resolution, with excellent agreement between observed and calculated structure factors (R_{cryst} = 9.85% for all 156,181 reflections, see Methods). The data extend to 0.90 Å resolution, and in fact one of the larger crystals showed reflections near 0.80 Å resolution. However, the data set is incomplete within the 0.95-0.90 Å resolution bin, so that the resolution should be considered 0.95 Å. Nevertheless, the quality of the data is unprecedented given the 35 kDa size of the protein, and a more detailed structural analysis will be provided in the 'Discussion' (Chapter 4). Crystals of PP8-penicillopepsin are also exceptional, and future studies will no doubt involve

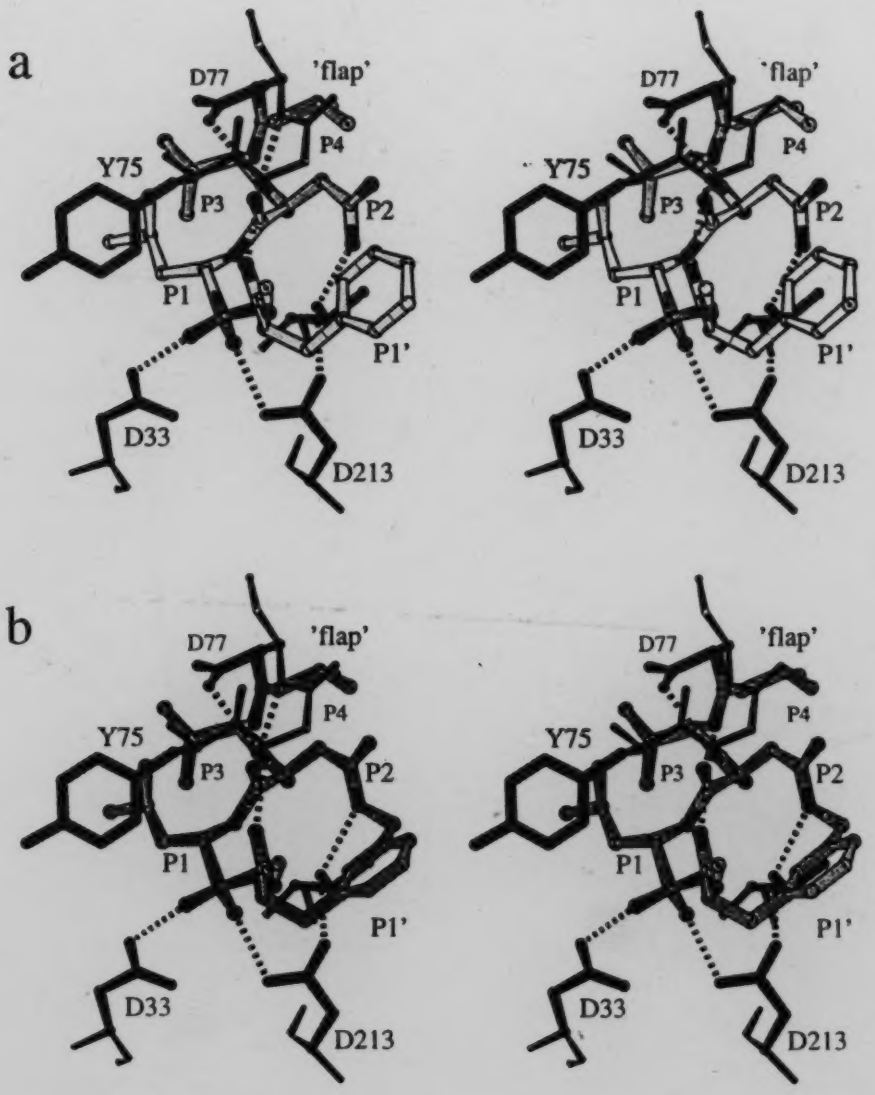


Figure 2C.2: Ball-and-stick models of PP7 and PP8 at the active site of penicillopepsin, shown in divergent stereo. (a) Structure of PP8-penicillopepsin. The inhibitor is distinguished by the lightly-shaded bond vectors. Hydrogen bonds are shown as dashed lines. (b) Structure of PP7-penicillopepsin. In this figure, the inhibitor is distnguished by the darker shading of bond vectors, relative to the enzyme. The hydrogen bonds are preserved in this complex, but the P2 and P1' side chains are bridged by an amide bond and methylene carbon atom.

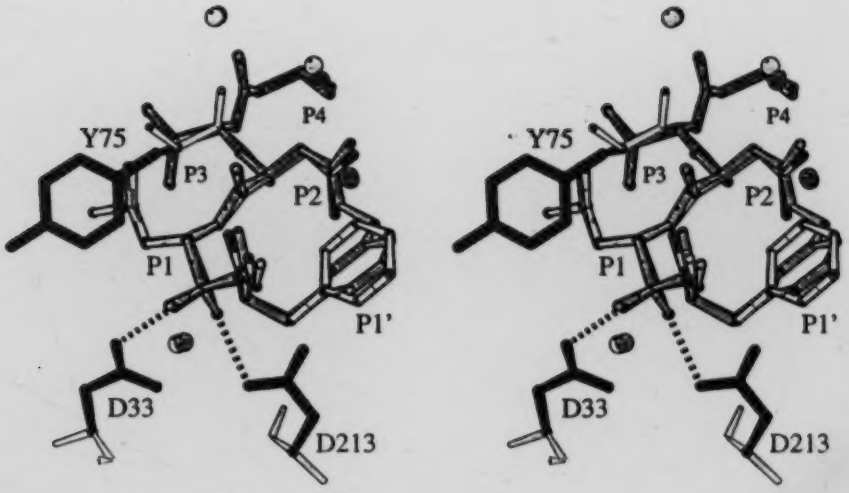


Figure 2C.3: Superpositions of PP7 and PP8 at the active site of penicillopepsin. The superposition of PP7 and PP8 was performed using the backbone enzyme atoms. PP7 is represented by the darker shading. Four conserved water molecules are represented by spheres. Thr216 and most of the flap residues have been excluded to clearly show the similarities among the two complexes. The rotation of the phenyl ring at P1' is apparent in the figure.

characterization of this complex beyond 1.0 Å resolution. Although many alternate conformations are observed for both PP7-penicillopepsin (30 residues) and PP8-penicillopepsin (12 residues), it is important to note that they occur in regions remote from the active site and do not affect the discussions below.

The position and conformation of PP7 in the complex with penicillopepsin is virtually identical to PP8 (Figs. 2C.2,3). Comparisons of the hydrogen bond distances, the surface areas buried, and the root-mean-square (rms) deviations of the complexes reveal very similar binding of these two inhibitors (Tables 2C.2-5). In addition, 11 of 13 water molecules in the vicinity of PP8 (< 4.0 Å) have structural equivalents in the complex of PP7-penicillopepsin. Upon superposition of the enzyme backbone, the rms deviation of these 11 waters is 0.31 Å.

The most pronounced conformational difference between PP7 and PP8 is a 36° rotation at the $\chi 2$-angle of P1'-Phe (Fig. 2C.3). In PP7-penicillopepsin, $\chi 2$ adopts a favorable value of -73° (Dunbrack & Karplus, 1994; Schrauber et al., 1993), whereas the $\chi 2$-value of PP8 is -37°. Rotation of the phenyl ring in PP8 is likely due to the close approach of P2-Asn. The side chain of P2-Asn is firmly held in its position by a hydrogen bond between its Nδ2 atom and Thr216(Oγ1) in both PP7 and PP8. In the cyclic PP7 structure, the P1'-Phe(Cε1) atom is situated 2.52 Å from Nδ2 of P2-Asn because of the methylene group that binds them together; removal of the methylene group to form acyclic PP8 would lead to a severe steric clash were the phenyl ring not rotated slightly, as observed in Fig. 2C.3a. The overall conformational change is also attributed to a relative 12° rotation of the phsophoester linkage between P1 and P1', such that the largest co-ordinate difference is 1.29 Å (Cε1 of P1'-Phe) upon superposition of the penicillopepsin backbone. The positions of the atoms of penicillopepsin that form the substrate clefts in the two complexes are identical in spite of the conformational difference observed at P1'-Phe.

The ω-angle corresponding to the amide bridge linking P2-Asn and P1'Phe is 168.6°, which is significantly distorted from an ideal value of 180°. The distortion from ideality is likely attributed to the favourable hydrogen bond between the Nδ2 atom of P2-Asn and the Oγ1 atom of Thr216. Such a deviation from planarity corresponds to a cost of about 1.0 kcal/mol (MacArthur & Thornton, 1996) and is an important consideration when tabulating the energetics of the binding process (see Discussion, below).

The conformation of PP7 in the complex is similar to the solution structure that was determined by a combined nuclear magnetic resonance (NMR) and modeling approach (Smith & Bartlett, 1998), suggesting that the macrocycle is not significantly distorted upon formation of the complex (Fig. 2C.4). However, an interesting structural difference is observed in the 'puckering' of the macocycle. The structural change is attributed to the geometry of the amide bridge, which is flipped 180° in the NMR model relative to the crystal structure. Least-squares superposition of all 44 common atoms of PP7 gives an rms deviation of 1.19 Å between the two structures. Neglecting the 10 atoms that comprise the phenyl ring and amide bridge, the rms deviation is reduced to 0.88 Å.

It is not evident why the orientation of the amide bridge is different in the structures. The cyclization step was performed subsequent to the assembly of the peptide backbone during synthesis of PP7 (Fig. 2C.5; Smith & Bartlett, 1998). In principle, the peptide coupling step should result in the formation of both geometric forms of PP7. However, there is no evidence from the electron density maps that the inhibitor with the peptide-bond orientation observed in the NMR structure is bound at the active site of penicillopepsin. Upon modeling of the NMR structure into the active site of PP7-penicillopepsin, the Oδ1 atom of P2-Asn resides only 1.4 Å away from the Oγ1 atom of Thr216. Therefore, it is possible that this geometric isomer of PP7 has a lower binding affinity and is unable to compete for the enzyme active site. Our

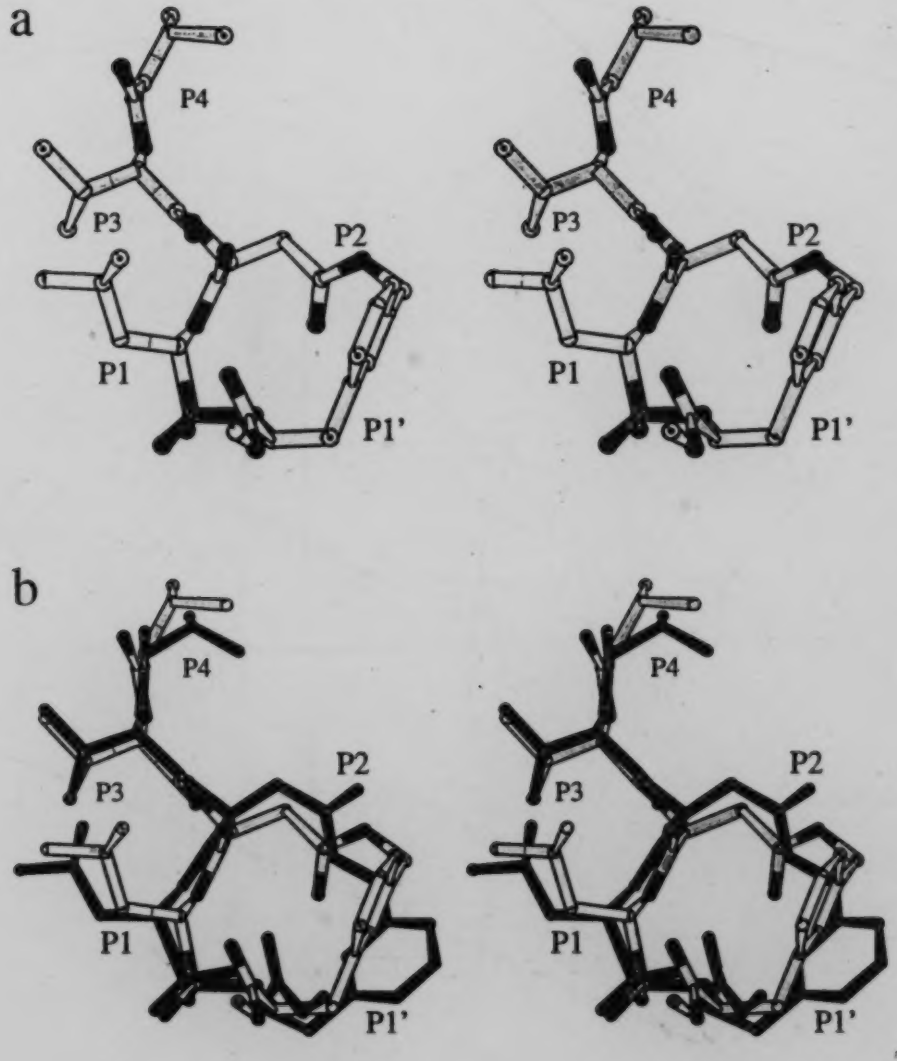


Figure 2C.4: NMR solution model of PP7. (a) Ball and stick representation of the solution structure derived from NMR and modeling studies (Smith & Bartlett, 1998). (b) Superposition of the NMR and crystal structures of PP7. The atoms comprising the P1' phenyl ring and the amide bridge were excluded from the calculation. If the NMR model were placed into the active site of penicillopepsin, the Oδ1 atom of P2-Asn would protrude down to within 1.4 Å of the Oγ1 atom of Thr216 (not shown), thus resulting in severe steric clashes.

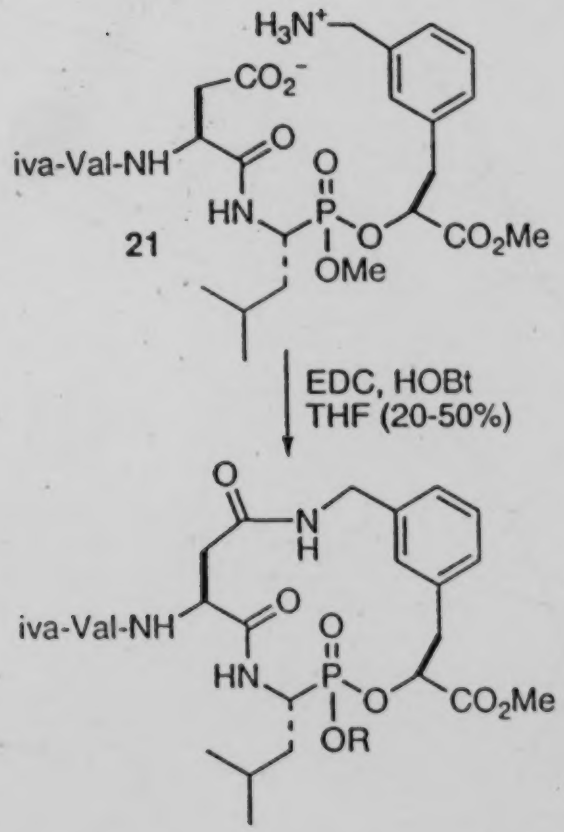


Figure 2C.5: The synthetic step leading to cyclization of PP7. The figure was adapted from Smith & Bartlett (1998). The reagents are ethyl dimethylaminoethyl carbodiimide (EDC) and hydroxybenzotriazole (HOBt). This peptide coupling step was performed in a dilute solution of tetrahydrofuran (THF), giving a yield of 20-50% for the reaction. The methyl substituent 'R' of the phosphonate methyl ester was removed in the subsequent step to form the final product, PP7.

collaborators at the University of California (Berkeley) are re-investigating the solution structure of PP7 in light of the crystal structure in order to determine whether the NMR data are consistent with the alternative peptide bond geometry.

Evaluation of the rational design strategy. The structures of PP7- and PP8-penicillopepsin can be compared to the 1.8 Å structure of PP2-penicillopepsin (K_i = 2.8 nM), the parent complex from which the inhibitors were designed (Fig. 2C.6; Tables 2C.2-5). This is an important analysis since it permits an evaluation of the extent to which the rational design of the macrocycle PP7 was successful. As expected from the design strategy, the conformations of the PP7 and PP8, their chemical interactions with the enzyme, and the overall surface area buried in their complexes with penicillopepsin are virtually identical to PP2-penicillopepsin. The water molecule that bridges the carbonyl oxygen of P3-Val to the backbone nitrogen of Asn218 is conserved in all three complexes. However, PP2 is more hydrophobic as a consequence of the valine residue at P2. The less intrusive sidechain of P2-Val also permits the χ2-angle of P1'-Phe to adopt a favorable value of -82° without steric clashes. The close structural similarities among the three enzyme-inhibitor complexes indicate that the attempt at rational design of macrocyclic PP7 was successful. The 15-fold lower K_i of PP2 relative to PP8, both of which are acyclic, may reflect the preference of the hydrophobic S2 pocket for P2-Val, thus contributing only to enthalpic differences.

The water structure in the vicinity of the active site reveals the most variability when comparing PP2-penicillopepsin to the second-generation inhibitor complexes (Fig. 2C.6). However, it is important to note that the programs used in refinement and the resolution of the data sets are quite different. Nevertheless, the water molecules tend to cluster in distinct areas, such as the region near the P2 and P1' side chains, which are more solvent exposed relative to the P1 and P3 side chains.

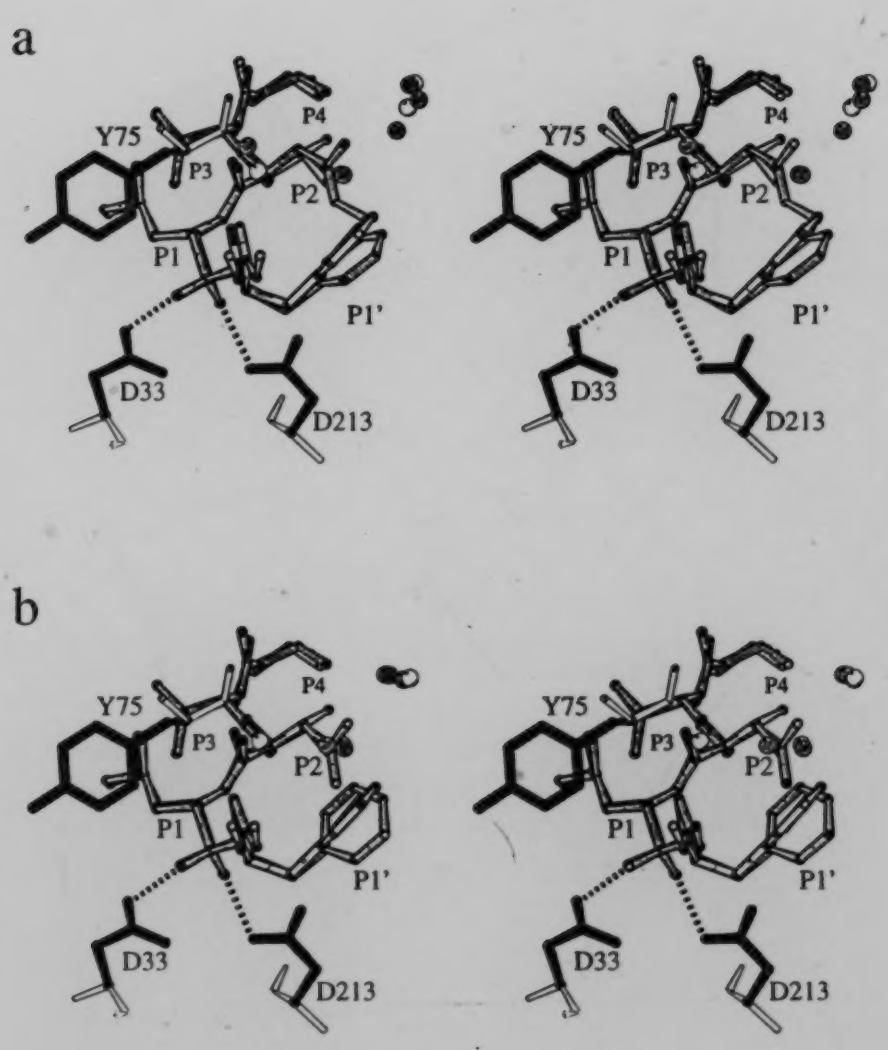


Figure 2C.6: Evaluation of the rational design strategy. (a) Superposition of PP2-penicillopepsin (light shading) with PP7-penicillopepsin (dark). Water molecules are shown as spheres. The water molecules in this region are mostly conserved in the two structures. The darkly shaded water in PP7-penicillopepsin that lies behind the P2 side chain obscures the conserved lightly shaded water of PP2. (b) Superposition of PP2-penicillopepsin (light shading) with PP8-penicillopepsin (dark). These models show the greatest variability in their active site water structures.

Table 2C.2: Possible Hydrogen-bonding Contact Distances Between Phosphonate Inhibitors and the Active Site

			distances (Å)		
site	inhibitor	penicillopepsin	PP2	PP7	PP8
P3	CO	Thr217 (NH)	3.03	3.04	3.04
	NH	Thr217 (Oγ1)	2.88	2.94	3.15
P2	NH	Asp77 (Oδ1)	2.93	2.89	2.83
	Nδ2 (Asn)	Thr216 (Oγ1)	--	3.02	2.97
	CO	Asp77 (NH)	3.00	3.14	3.09
P1	NH	Gly215 (O)	3.22	3.06	3.03
	POH	Asp33 (Oδ1)	2.40	2.42	2.39
	PO	Asp33 (Oδ2)	3.21	3.17	3.16
	PO	Asp213 (Oδ1)	2.60	2.54	2.51
P1'	CO (ester)	Gly76 (NH)	2.89	2.92	2.95

Conformational χ2-angles for P1'-Phe

PP2 = -82° PP7 = -73° PP8 = -37°

†Largest deviations between PP7 and PP8

Cδ of Phe-P1': Δ = 1.15 Å
Methyl-group of ester: Δ = 1.10 Å

†These distances were determined following least-squares superposition of the backbone atoms of the enzyme (excluding the inhibitors).

Table 2C.3: Solvent-Accessible Surface Area ($Å^2$)†

inhibitor¥	uncomplexed			complexed			% buried
	hydrophilic	hydrophobic	total	hydrophilic	hydrophobic	total	
§PP2	192	673	866	26	124	150	83
PP7	245	636	881	40	121	161	82
PP8	256	628	880	38	123	161	82

†Calulated using the algorithm of Richards (Richards, 1977), as implemented by the program of Connolly (Connolly, 1983a/b).

¥The PDB files were first stripped of water molecules.

§PDB code 1ppl (Fraser et al., 1992).

Table 2C.4: Conformational angles (deg) of Phosphorus-Based Peptide Inhibitors Bound to Penicillopepsin

site	angle	PP2	PP7	PP8
P4	ψ*	-140	-122	-121
	ω*	-174	-176	-172
	χ1*	59	52	55
P3	ϕ	-123	-125	-131
	ψ	162	162	160
	ω	179	-175	-174
	χ1	-57	-58	-58
P2	ϕ	-132	-142	-139
	ψ	99	82	81
	ω	-180	-172	-175
	χ1	-176	-177	177
	χ2	-----	-153	-148
P1	ϕ	-123	-116	-108
	ψ	73	72	64
	ω	147	147	142
	χ1	-48	-48	-59
	χ2	159	175	160
P1'	ϕ	-113	-127	-113
	ψ	174	-175	-169
	χ1	63	67	73
	χ2	-83	-73	-38

Torsion angle definitions for isovaleryl at P4:

ψ* = Cβ-Cα-C-N

ω* = Cα(Iva)-C-N-Cα

χ1* = Cγ1-Cβ-Cα-C

Table 2C.5. Least-Squares Superpositions of Enzyme-Inhibitor Complexes

	PP2	PP7	PP8
PP2	–	0.28 (1260)	0.30 (1260)
PP7	0.30 (34)	–	0.19 (1260)
PP8	0.25 (34)	0.17 (37)	–

The values given are rms deviations in Å units. The upper right portion of the matrix is a superposition of the enzyme backbone. Residues 1-2, 278-281, and 322-323 were disordered and therefore excluded from the calculations. The bottom left region of the matrix is a superposition of the inhibitor atoms. The P1'-Phe ring atoms were excluded from these calculations. The numbers in brackets indicate the number of common atoms used in the superposition.

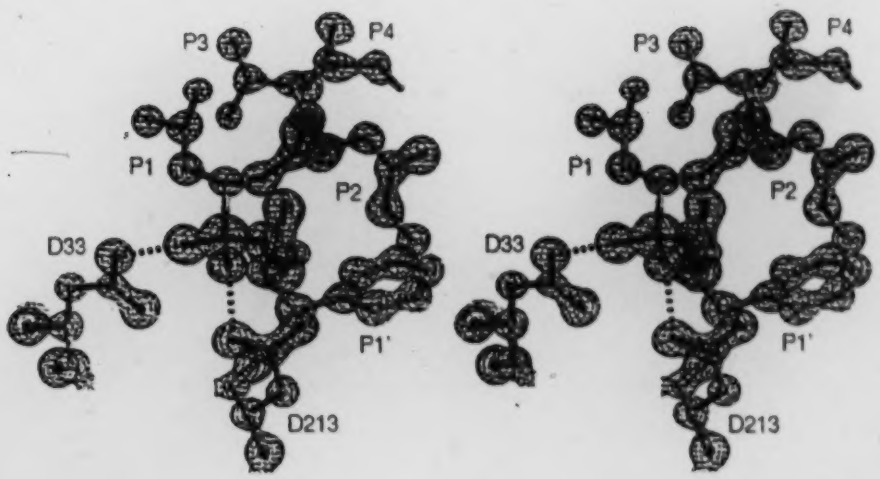


Figure 2C.7: Electron density at the active site of PP7-penicillopepsin. The refined $2F_o$-F_c map has been contoured at 2.5σ, revealing the density around individual atom positions.

D. DISCUSSION

There are many enthalpic and entropic contributions to the binding of a peptide to a protein. For example, the entropic terms include the desolvation and rigidification of those residues in the ligand and the protein that mediate formation of the complex, as well as the loss of global rotational and translational degrees of freedom in the bimolecular association. In the complexes of PP7- and PP8-penicillopepsin, the similarities in the chemical interactions (hydrogen bonds, van der Waals contacts) and the amount of surface areas buried (82%) suggest that many of these energy terms are almost identical. There is only the slight movement of the phenyl ring at P1'-Phe that could represent a small enthalpic difference in their binding energies. It is apparent that the 420-fold difference in binding affinity between inhibitors PP7 and PP8 can mostly be attributed to the entropy that is associated with internal bond rotations. The overall free energy is given by,

$$\Delta\Delta G^{\circ} = -RT[\ln (K_{i,cyc}/K_{i,acy})] \qquad (1)$$

where R is the universal gas constant, T is the temperature (in Kelvin) and K_i is the inhibition constant (moles/litre) for the cyclic and acyclic inhibitors. The $\Delta\Delta G^{\circ}$ value of 3.6 kcal/mol derived from equation 1 represents the entropic penalty required to freeze out the enzyme-bound conformation from the flexible conformations of PP8, relative to PP7, during the binding process.

The entropic cost of conformational mobility can be estimated on a per-bond basis, as described previously (Smith & Bartlett, 1998). Briefly, a linear chain of *n*-atoms has (*n*-1) independent rotations. When the termini are linked, an additional rotatable bond is added, but 6 degrees of freedom are lost because the bond rotations become interdependent. In the acyclic analogue PP8, there are 9 single, rotatable bonds between the P2 side chain carboxamide and the P1' phenyl group; within the

macrocycle of PP8, there are 11 single, rotatable bonds, corresponding to 5 degrees of freedom. The 3.6 kcal/mol difference in free energy of binding for these two inhibitors thus corresponds to an average of 0.9 kcal/mol per degree of freedom due to bond rotations. This experimental value is well within the range of 0.7-1.6 kcal/mol reported in the literature from studies of small molecules and the thermodynamic behaviour of helix-forming peptides (Gomez & Freire, 1995; Williams et al., 1993; Page & Jencks, 1971; Andrews et al., 1984). The distortion from planarity of the amide bridge linking the P2 and P1' side chains is an additional penalty that must be overcome during the binding process. Thus, it may have the effect of underestimating the value of 0.90 kcal/mol attributed to each bond rotamer.

As described in the introduction to Chapter 2, the structure of the parent PP2-penicillopepsin complex suggested the feasibility of linking the side chains of P3-Val and P1-Leu. Accordingly, a macrocylic phosphonate inhibitor was synthesized in which the P3 and P1 side chains were linked by a naphthalene bridge. The structure of its complex with penicillopepsin was determined along with those of two acyclic analogues, the P1- and P3-naphthyl derivatives (PP4, PP5, and PP6; Ding et al., 1998). The binding affinities within this series of inhibitors increased as the degree of conformational freedom decreased, suggesting that the overall strategy was successful. However, the structures of the enzyme-inhibitor complexes were sufficiently different to prevent assignment of the free energy entirely to the conformational flexibility of the inhibitors. In this respect, the success of the inhibitors described here may be related to the more peripheral location and less intrusive nature of the amide bridge linking the P2 and P1' side chains (Fig. 2D.1).

Concluding Remarks. The strategy employed here provides the most rigorous and successful attempt to differentiate the entropy of internal bond rotations from the remaining free energy binding parameters in a large and complex system. Clearly,

further structural and thermodynamic studies would be required for more detailed analyses, such as a residue-specific entropy dependence. Such quantitative studies are a necessary step toward a better understanding of the forces that drive molecular recognition and accurate predictions of binding affinities (Searle & Williams, 1992). Recently, macrocyclic inhibitors have been synthesized against proteinases to enhance their potency and selectivity, but without an accompanying quantitative analysis of the effect [*e.g.*, renin, (Weber et al., 1992); and HIV proteinase (Ettmayer et al., 1996)]. The large 420-fold increase in the binding affinity observed here provides justification for pursuing this strategy in drug design, and is widely applicable to enzyme-inhibitor complexes.

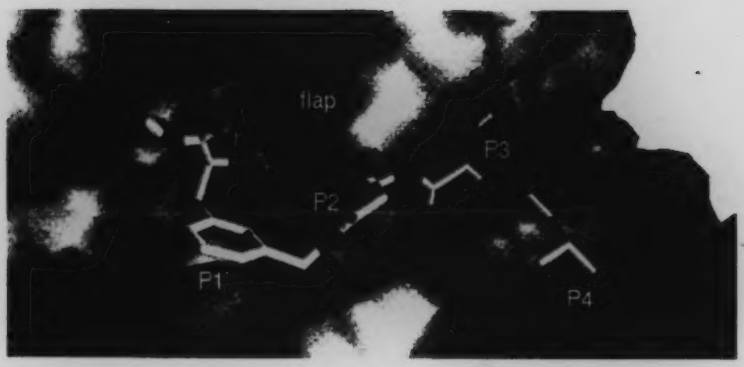


Figure 2D.1: Stick model of PP7 against the molecular surface of penicillopepsin. The covalent link between P2 and P1' is more peripherally located relative to the P3-to-P1 bridge that has been studied previously. The latter would face toward the enzyme pocket underneath the flap region of penicillopepsin.

E. REFERENCES

Andrews, P.R., Craik, D.J. & Martin, J.L. (1984) Functional group contributions to drug-receptor interactions. *J Med Chem* **27**: 1648-1657.

Baker, E.N. & Hubbard, R.E. (1984) Hydrogen bonding in globular proteins. *Prog Bioph Mol Biol* **44**: 97-179.

Bartlett, P.A., Hanson, J.E., Giannousis, P.P. (1990) Potent inhibition of pepsin and penicillopepsin by phosphorus-conaining peptide analogs. *J Org Chem* **55**: 6268-6274.

Blundell, T.L. et al. (1986). On the rational design of renin inhibitors: X-ray studies of aspartic proteinases complexed with transition-state analogues. *Biochemistry* **26**: 5585-5590.

Brunger, A.T. (1993) "XPLOR: a system for X-ray crystallography and NMR", Yale University Press, New Haven, CT.

Connolly, M.L. (1983a) Solvent-accessible surfaces of proteins and nucleic acids. *Science* **221**: 709-713.

Connolly, M.L. (1983b) Analytical molecular surface calculation. *J Appl Crystallogr* **16**: 548-558.

Davies, D.R. (1990) The structure and function of the aspartic proteinases. *Ann Rev Bioph Bioph Chem* **19**: 189-215.

Ding, J., Fraser, M.E., Meyer, J.H., Bartlett, P.A. & James, M.N.G. (1998) Macrocyclic inhibitors of penicillopepsin. 2. X-ray crystallographic analyses of penicillopepsin complexed with a P3-P1 macrocyclic peptidyl inhibitor and with its two acyclic analogues. *J Am Chem Soc* **120**: 4610-4621.

Dunbrack, R.L. & Karplus, M. (1994) Conformational analysis of the backbone-dependent rotamer preferences of protein sidechains. *Nature Struct Biol* **1**: 334-339.

Engh, R.A. & Huber, R. (1991) Accurate bond and angle parameters for X-ray protein structure refinement. *Acta Cryst* **A47**: 392-400.

Esnouf, R.M. (1997) An extensively modified version of MolScript that includes greatly enhanced colouring capabilities. *J Mol Graphics* **15**: 133-138.

Ettmayer, P., Billich, A., Hecht, P., Rosenwirth, B. & Gstach, H. (1996) Paracyclophanes: A novel class of water-soluble inhibitors of HIV proteinase. *J Med Chem* **39**: 3291-3299.

Evans, P.R. (1993) Data Reduction, In *Proceedings of the CCP4 Study Weekend.* pp. 114-122.

Fraser, M.E., Strynadka, N.C.J., Bartlett, P.A., Hanson, J.E. & James, M.N.G. (1992) Crystallographic analysis of transition-state mimics bound to penicillopepsin: Phosphorus-containing peptide analogues. *Biochemistry* **31**: 5201-5214.

French, G.S. & Wilson, K.S. (1978) On the treatment of negative intensity observations. *Acta Cryst* **A34**: 517-525.

Gomez, J. & Freire, E. (1995) Thermodynamic mapping of the inhibitor site of the aspartic protease endothiapepsin. *J Mol Biol* **252**: 337-350.

Holden, H.M., Tronrud, D.E., Monzingo, A.F., Weaver, L.H. & Matthews, B.W. (1987) Slow- and fast-binding inhibitors of thermolysin display different modes of binding: crystallographic anlalysis of extended phosphonamidate transition-state analogues. *Biochemistry* **26**: 8542-8553.

Howard, A.J. et al. (1987) The use of an imaging proportional counter in macromolecular crystallography. *J Appl Cryst* **20**: 383-387.

James, M.N.G., Sielecki, A.R., Hayakawa, K. & Gelb, M.H. (1992) Crystallographic analysis of transition state mimics bound to penicillopepsin: difluorostatine- and difluorostatone-containing peptides. *Biochemistry* **31**: 3872-3886.

Jones, T.A. & Kjeldgaard, M. (1995) O - the Manual, Version 5.11 (Uppsala, Sweden).

Kim H, Lipscomb WN. 1991. Comparison of the structures of three carboxypeptidase A-phosphonate complexes determined by X-ray crystallography. *Biochemistry 30*: 8171-8180.

Lauri, G. & Bartlett, P.A. (1994) CAVEAT: a program to facilitate the design of organic molecules. *J Comp Aided Mol Design* **8**: 51-66.

Leslie, A.G.W. MOSFLM Version 5.50, MRC Laboratory of Molecular Biology, University of Cambridge, United Kingdom.

Luzzati, V. (1952) Traitement statistique des erreurs dans la determination des structures cristallines. *Acta Crystallogr* **5**: 802-810.

MacArthur, M.W. & Thornton, J.M. (1996) Deviations from planarity of the peptide bond in peptides and proteins. *J Mol Biol* **264**: 1180-1195.

McCree, D.E. (1992) A visual protein crystallographic software for X11/XView. *J Mol Graphics* **10**: 44-46.

Meyer, J.H. & Bartlett, P.A. (1998) Macrocyclic inhibitors of penicillopepsin. 1. Design, synthesis, and evaluation of an inhibitor bridged between P1 and P3. *J Am Chem Soc* **120**: 4600-4609.

Nicholls, A., Sharp, K.A. & Honing, B. (1991) Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins* **11**: 281-296.

Page, M.I. & Jencks, W.P. (1971) Entropic contributions to rate accelerations in enzymic and intramolecular reactions and the chelate effect. *Proc Natl Acad Sci (USA)* **68**: 1678-1683.

Read, R.J. (1986) Improved fourier co-efficients for maps using phases from partial structures with errors. *Acta Cryst* **A42**: 140-149.

Richards, F.M. (1977) Areas, volume, packing, and protein structure. *Ann Rev Biophys Bioeng* **6**: 151-176.

Salemme, F.R., Spurlino, J. & Bone, R. (1997) Serendipity meets precision: the integration of structure-based drug design and combinatorial chemistry for efficient drug discovery. *Structure* **5**: 319-324.

Schrauber, H., Eisenhaber, F. & Argos, P. (1993) Rotamers: to be or not to be? An analysis of amino acid side-chain conformations in globular proteins. *J Mol Biol* **230**, 592-612.

Searle, M.S. & Williams, D.H. (1992) The cost of conformational order: entropy changes in molecular associations. *J Am Chem Soc* **114**: 10690-10697.

Sheldrick, G.M. (1997) The SHELX-97 Manual, University of Gottingen, Germany.

Smith, W.W. & Bartlett, P.A. (1998) Macrocyclic inhibitors of penicillopepsin. 3. Design, synthesis and evaluation of an inhibitor bridged between P2 and P1'. *J Am Chem Soc* **120**: 4622-4628.

Weber, A.E., Steiner, M.G., Krieter, P.A., Colletti, A.E., Tata, J.R., Halgren, T.A., Ball, R.G., Doyle, J.J., Schorn, T.W. & Stearns, R.A. (1992) Highly potent, orally active diester macrocyclic human renin inhibitors. *J Med Chem* **35**: 3755-3773.

Williams, D.H., Searle, M.S., Mackay, J.P., Gerhard, U. & Maplestone, R.A. (1993) Toward an estimation of binding constants in aqueous solution: studies of associations of vanomycin group antibiotics. *Proc Natl Acad Sci (USA)* **90**: 1172-1178.

CHAPTER 3

STRUCTURAL CHARACTERIZATION OF ACTIVATION 'INTERMEDIATE 2' ON THE PATHWAY TO HUMAN GASTRICSIN

A. INTRODUCTION

All known proteolytic enzymes in nature are synthesized as inactive precursors to prevent unwanted protein degradation within cells. As discussed previously, the gastric aspartic proteinases are synthesized in the chief cells of the stomach as zymogens and contain an inhibitory N-terminal prosegment of 43-47 residues. The conversion process is autocatalytic at low pH, and has been studied extensively using biochemical, kinetic and structural techniques.

Structure of human progastricsin. The crystal structure of human progastricsin (hPGC; Moore et al., 1995), the second zymogen characterized in the aspartic proteinase family after pig pepsinogen (pPGN), revealed that the secondary structure of the prosegment and the overall mechanism of inhibition is identical for these two zymogens (Figs. 3A.1-3). The N-terminal β-strand and the three α-helices in the prosegment are conserved, while the largest difference is confined to the loop region (R39p-E5) following helix 3p. This loop contains the peptide bond that is eventually cleaved during maturation of the zymogen to the active enzyme. In hPGC, the path of this loop winds back through the active site cleft, while the corresponding loop in pPGN leads away from the cleft. Following the loop that contains the pro-mature junction, the paths of hPGC and pPGN converge at Pro6 of the mature segment. As a consequence of the intrusion of this region in hPGC, the "flap" (residues 71-81) and an adjoining loop (residues 125-134) are shifted away from the active site relative to the equivalent loops in pPGN.

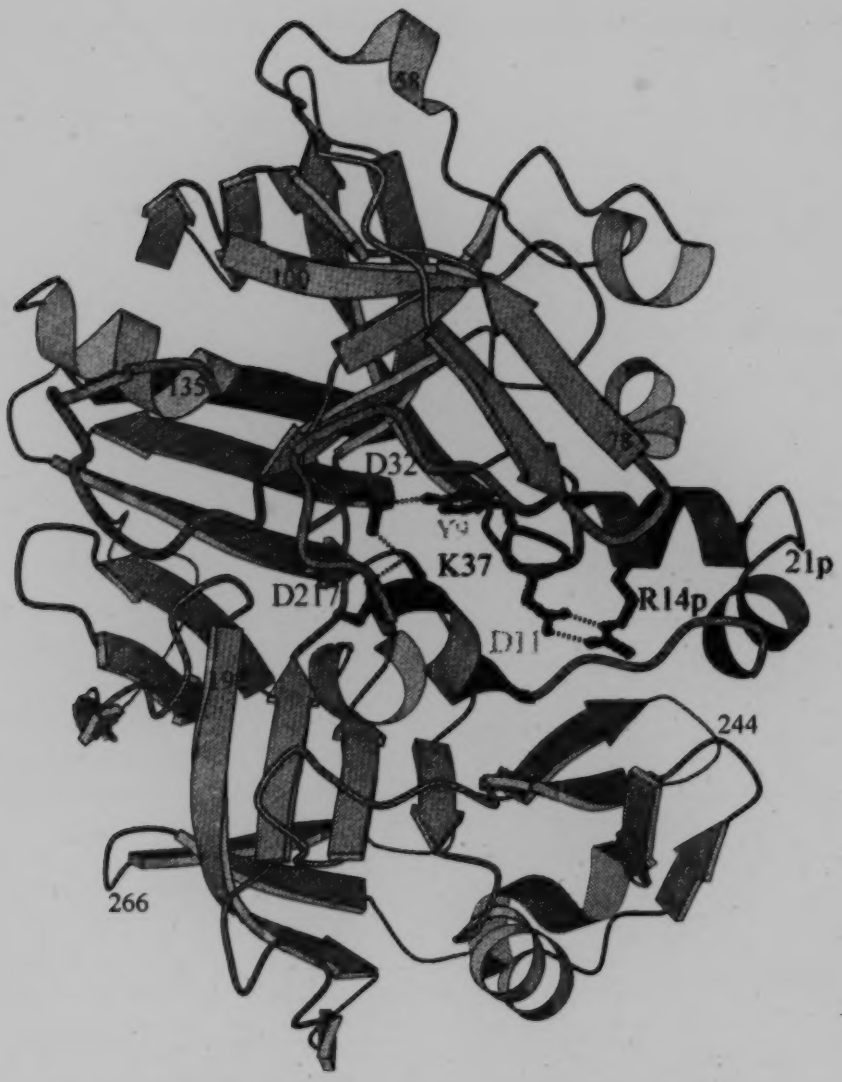


Figure 3A.1: Ribbon model of human prosgastricsin (Moore et al., 1995). The prosegment is green, and the regions of the mature segment that undergo conformational changes during conversion are pink. Hydrogen bonds are shown as dashed lines. Lys37p in helix 3p of the prosegment makes salt-bridges to the active site aspartate residues (Asp32, Asp217), thereby positioning the inhibitory part of the prosegment in front of the active site and preventing access to substrates.

Figure 3A.2: Superposition of human progastricsin with porcine pepsinogen, shown as C_α vectors in divergent stereo. Progastricsin is shown as solid lines, while pepsinogen is represented by dashed lines. The prosegment of progastricsin is emphasized with thick lines. The zymogens were aligned as previously described (Moore et al., 1995). The rms difference for 1076 common main chain atoms is 1.5 Å.

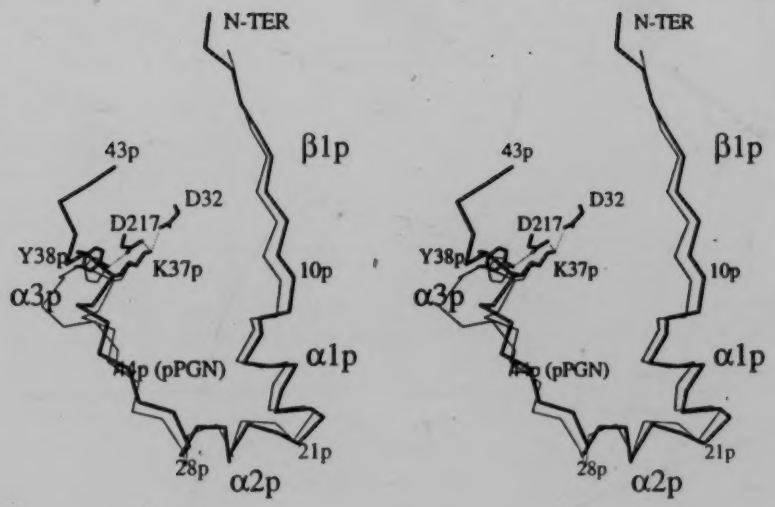


Figure 3A.3: Superposition of the prosegments of human progastricsin and porcine pepsinogen, shown as C_α vectors in stereo. Thick lines correspond to progastricsin; thin lines represent pepsinogen. Hydrogen bonds are shown with dashed lines. The "p" suffix refers to those residues that reside in the prosegment. The secondary structure elements (β1p, α1p-α3p) are conserved between progastricsin and pepsinogen. Following α3p, the main chains follow a very diffferent course leading to the pro-mature junction.

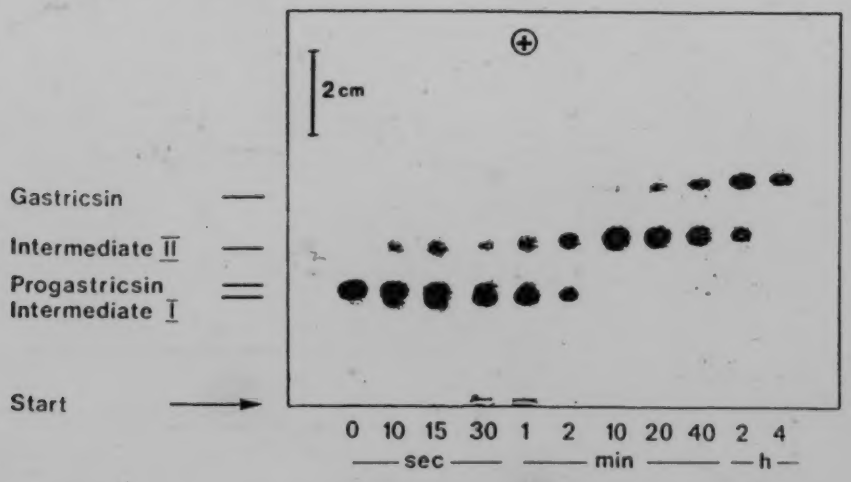


Figure 3A.4: Analyses of the activation pathway for human progastricsin, performed by Foltmann and Jensen (1982). The zymogen was incubated at pH 2 and O°C. Samples were analyzed at various times using non-denaturing agarose gel electrophoresis. 'Intermediate 1' migrates slower than the zymogen early in the activation pathway, although no cleavages have taken place.

The structure of hPGC reveals that a Lys residue (Lys37p) in the prosegment forms salt bridges with the active site aspartate residues (Asp32 and Asp217), thereby positioning the prosegment in front of the active site and preventing the approach of substrates. Secretion of the zymogen from the cytosol into an acidic medium (pH ~2.0) would likely result in the protonation of many aspartate and glutamate residues. Thus, the activation process is presumably initiated by a weakening of the salt bridges between Lys37p (helix 3p) and the active site Asp residues. A pair of 3_{10}-helices protrude into the active site cleft, one from the prosegment (α3p) and one from the mature segment (Pro6-Met10). Sidechain hydrogen bonds between Tyr38p--Asp217 and Tyr9--Asp32 help to stabilize the position of these two short helices in front of the active site (Fig. 3A.1,3)

The conversion process has been studied by Bent Foltmann using a variety of biochemical techniques. Upon incubation in acid, the initial intermediate of human progastricsin can be detected by non-denaturing agarose gel electrophoresis, and has been termed 'intermediate 1' (Foltmann & Jensen, 1982). These classical experiments were performed on ice in order to slow the conversion process from seconds to several minutes, thus enabling the identification of activation intermediates (Fig. 3A.4). The transition from hPGC→intermediate 1 can be reversed by rapidly returning the pH of the solution to neutrality (Foltmann & Jensen, 1982). It is likely that the helical regions of the prosegment become unravelled in intermediate 1, thus resulting in altered migration of the protein upon gel electrophoresis under non-denaturing conditions (Moore et al., 1995). Upon further incubation of progastricsin in acid, a more stable protein species ('intermediate 2'; Fig. 3A.4) is always observed prior to the appearance of mature gastricsin (Foltmann & Jensen, 1982; Nielsen & Foltmann, 1993). Conversion is predicted to involve partial removal of the prosegment in the initial stages of activation, and this hypothesis is consistent with the failure to isolate the full-length (43-residue) prosegment *in vitro*.

The enzyme species with altered migrations have been carefully analyzed by various techniques including amino acid analysis and sequencing, chromatography, and mass spectroscopy (Foltmann & Jensen, 1982; Nielsen & Foltmann, 1993). These studies have revealed that the first hydrolytic event during activation of hPGC is the uni-molecular cleavage of the peptide bond between Phe26p and Leu27p (Foltmann & Jensen, 1982). Subsequently, intermolecular cleavage at the pro-mature junction (Leu43p-Ser1) results in a transient intermediate (intermediate 2, or hGSi) in which residues Ala1p-Phe26p are non-covalently associated with the mature enzyme (Ser1-Ala329), while the remainder of the prosegment (Leu27p-Leu43p) is removed. The intermediate complex can be stabilized by transfer to neutral pH, thereby preventing further activation events and inhibiting proteolytic activity (Foltmann & Jensen, 1982). The nature of the interaction between the prosegment peptide and mature gastricsin in hGSi, and the mechanism of inhibition, have not been determined previously.

These observations are interesting when comparing the activation pathways of progastricsin and pepsinogen. Stepwise cleavages within the prosegment also take place during acid activation of porcine pepsinogen to form the equivalent 'intermediate 2'. In this case, the initial bond cleavage takes place between Leu16p and Leu17p, resulting in a transient intermediate that has been termed 'pseudopepsin' (residues 17p-326; Dykes & Kay, 1976; Al-Janabi et al., 1972). However, the full-length prosegment of pepsinogen (1p-44p) can also be isolated during conversion (Kageyama & Takahashi, 1984), suggesting that multiple and concurrent pathways lead to the mature enzyme. As described below, the ability to isolate a discrete intermediate of progastricsin has allowed a structural characterization of the activation pathway of aspartic proteinases for the first time.

In summary, zymogens from the gastric aspartic proteinases share sequence and structural similarities in their prosegments. Consequently, the mechanisms by which the zymogens are converted to active enzymes at low pH are likely to share

common features. However, the detailed activation pathways and the kinetics of the conversion process are variable, as evidenced by studies of progastricsin and pepsinogen. Here, the structure of 'intermediate 2' of human progastricsin has been determined with a view to understanding the molecular details and biological significance of the conversion process. Despite the variations in the kinetics of activation and the identities of intermediate species, it will be argued that some essential features of the activation pathway are conserved among all gastric aspartic proteinases.

B. MATERIALS AND METHODS

Purification and crystallization. The intermediate was prepared by Dr. N. Tarasova from human progastricsin by auto-activation, essentially as described by Bent Foltmann (Foltmann & Jensen, 1982), and is termed "Intermediate 2" in his paper. Briefly, human progastricsin was extracted from gastric mucosa at pH 6.3. Auto-activation of the zymogen was performed at pH 2.0 for 20 min, and subsequently all solutions were maintained above pH 5.0. The intermediate was initially fractionated from the crude extract by anion exchange and gel filtration chromatography, purified by FPLC, and was >99% pure as judged by SDS-polyacrylamide gel electrophoresis and electro-spray mass spectrometry (data not shown). The intermediate was crystallized in 4M sodium formate, 100 mM Bis-Tris-propane (BTP), pH 7.8, using the hanging drop vapour diffusion method. Crystals grew in the monoclinic space group C2, with the unit cell dimensions shown in Table 3B.1.

Data collection and structure determination. Data were collected at 100K at the DESY in Hamburg, and processed using the program DENZO and Scalepack (Otwinowski, 1993). The structure was solved by AMoRe (Navaza, 1994) using the previously determined structure of hPGC as the search model (Moore et al., 1995). Residues Ala1p-Leu23p and Met7-Ala329 were used in the rotation and translation search. Two solutions were obtained and subjected to rigid body refinement in AMoRe. Refinement was performed using a maximum likelihood target function (Pannu & Read, 1996), implemented in X-PLOR (Brunger, 1993), using data between 20 Å and 2.36 Å ($F/\sigma F > 2.0$). Non-crystallographic symmetry restraints (in X-PLOR) and density averaging (AVERAGE in the CCP4 package, 1994) were used throughout model building and refinement. Manual intervention, display and map interpretation were performed using the program O (Version 5.11; Jones & Kjeldgaard, 1995). The final model includes Ala1p-Gly21p of the prosegment, Val2-Ala329 of mature gastricsin, 312 water molecules and one Na^+ ion. Electron density

Table 3B.1: Statistics of Data and Refinement

Space group	C2
Unit cell parameters	a=156.3Å b=50.4Å c=125.2Å β=117.6°
Completeness (20.0-2.36Å)	95.5%
Completeness (2.40-2.36Å)	80.8%
Multiplicity (20.0-2.36Å)	4.1
R_{merge}[1]	6.6%
$\langle I/\sigma(I)\rangle$ (20-2.36Å)	9.2
Unique reflections	34523
Reflections with F/sigF > 2.0	34124
Atoms in asymmetric unit	5207
Molecules in asymmetric unit	2
R-factor[2] (20-2.36Å)	22.4%
R-free[3]	27.1%

R.m.s. deviations from ideal geometry

Bond lengths (Å)	0.011
Bond angles (°)	1.8
Dihedral angles (°)	26.7

[1] $R_{merge} = \sum_{hklj}|I_{hklj}-\langle I_{hkl}\rangle|/\sum_{hklj}\langle I_{hkl}\rangle$

R-factors (below) were calculated using data with F/sigF > 2.0

[2] R-Factor = $\sum||F_o|-|F_c||/\sum|F_o|$ where $|F_o|$ and $|F_c|$ are observed and calculated structure factor amplitudes, respectively.

[3] R-free was calculated from a random set containing 10% of the observations that were omitted during refinement.

for residues Tyr75 and Gly76 was poor. There was no observed density for the loop Asn157-Ser162, which is poorly defined in all aspartic proteinase structures.

The refined structure of hGSi was superimposed with the zymogen, excluding the residues Ala124-Thr135 and Val2-Ala12, which had experienced major conformational changes (see succeeding pages). The rms displacement was 1.14 Å using a total of 1284 backbone atoms in the calculation.

C. RESULTS AND DISCUSSION

The overall structure of hGSi (Ala1p-Phe26p and mature gastricsin) consists of two β-sheet domains that are typical of the aspartic proteinase family (Fig. 3C.1). The catalytic Asp residues (Asp32, Asp217) are located on two loops at the base of a central hydrophobic groove, in front of a six-stranded anti-parallel β-sheet. Strikingly, the structure of hGSi reveals that the prosegment no longer obstructs access to the catalytic cleft, as it does in the zymogen. The excised prosegment (Leu27p-Leu43p) has vacated the active site region. A water molecule (Wat11) now resides between Asp32 and Asp217, characteristic of a mature enzyme containing a nucleophilic water that is poised for catalysis. Wat11 replaces the εNH_3^+ side chain of Lys37p that has been removed as part of the C-terminus of the prosegment during the formation of hGSi. However, the first 20 residues of the prosegment are held in the same position and conformation as was observed in the structure of the zymogen. This region consists of a β-strand (Val3p to Lys7p) that makes antiparallel β-sheet hydrogen bonds with a strand of the mature gastricsin moiety (Gly163-Phe168), and a subsequent α-helix (Ile13p to Lys20p; Fig. 3C.2). Asp11 of mature gastricsin forms a salt bridge with Arg14p, located in the helical prosegment of hGSi, as it does in hPGC.

In the zymogen structure, the loop Ala124-Thr135 is flipped away from the active site and has high temperature factors due to the proximity of the C-terminal region of the prosegment (Fig. 3A.1). In hGSi, this loop is ordered and closely resembles the corresponding region in human pepsin (Fujinaga et al., 1995), with a root-mean-square deviation of 0.6 Å for 48 backbone atoms. This loop is structurally conserved among all active aspartic proteinases; it had an unusual conformation in progastricsin. In the structure of hGSi, hydrogen bonds (not shown) are observed between Tyr125(O) and Ser36(Oγ), as well as Val130(O) and Asn37(Oδ1); these interactions are not observed in the zymogen. Ser36 and Asn37 are located on the reactive site loop that contains Asp32. A second loop that contains a β-hairpin (the

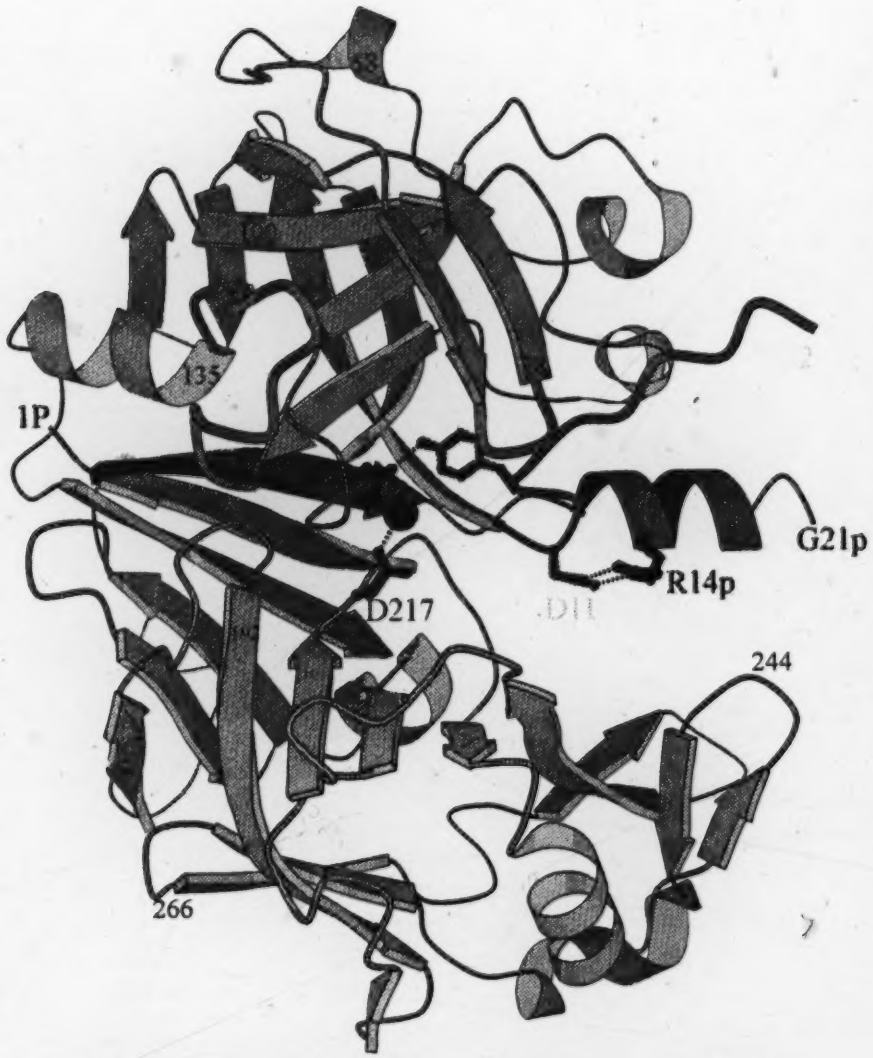


Figure 3C.1: Structure of 'intermediate 2' of human gastricsin, or hGSi. The remaining prosegment, 1p-21p, is green; the regions of the mature segment that undergo conformational changes (1-13, 72-80, and 124-135) are pink. The nucleophilic water molecule is a sphere and is a hallmark of active aspartic proteinases. Hydrogen bonds are represented by dashes.

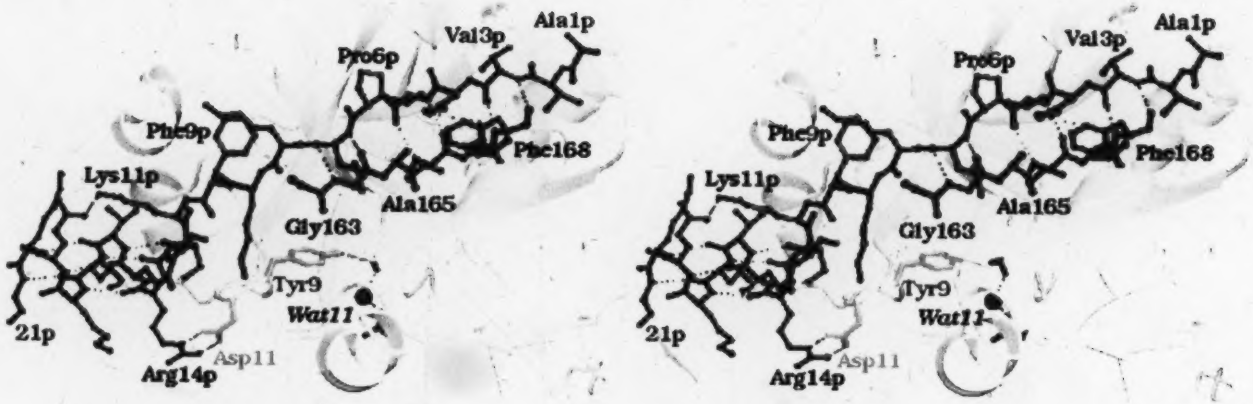


Figure 3C.2: Stick model of the prosegment of hGSi. The view is rotated by 180 degrees from the previous figure, providing a clear glimpse of the prosegment b-strand (in green). The partner strand of is coloured blue, and hydrogen bonds are dashed lines.:

"flap", Phe71-Thr81) also shifts toward the active site upon ejection of the C-terminal part of the prosegment. The flexibility of the flap region is apparent from the structures and the crystallographic B-factors of active aspartic proteinases. Structures of aspartic proteinases with bound inhibitors have shown that the flap closes to bind the inhibitor more tightly (Davies, 1990). The structure of the intermediate also demonstrates this flexibility during the activation process; the flap is "open" in the zymogen to accommodate the prosegment, and closes upon its removal in hGSi (Fig. 3C.3).

Mechanism of inhibition. The intermediate complex is inactive because Tyr9 of mature gastricsin forms a hydrogen bond with the carboxylate of Asp32 and partially blocks the S1 binding pocket, as it does in the zymogen (Fig. 3C.1,3). The first 8 residues of the mature segment in the zymogen consist of a loop (Ser1-Tyr4) and 3_{10}-helix (Glu5-Met10) that reside in the active site. However, the 3_{10}-helix is presumably disrupted during the first two cleavage events. Residues Val2-Ala8 have moved by up to 30 Å to form the intermediate, hGSi (Figs. 3C.3). This segment now has an extended conformation that is directed away from the active site, and is stabilized by symmetry contacts (Fig. 3C.4). The overall movement is the result of an uncoiling of the 3_{10}-helix, and a pivot preceding the Ala8-Tyr9 peptide bond. In the absence of crystallographic contacts, this region is likely to be flexible. During conversion, residues Ser1-Met10 eventually form the first β-strand as part of the six-stranded β-sheet common to all cellular aspartic proteinases, replacing residues Ala1p-Lys10p in the zymogen (Fig. 3C.5-7). However, the N-terminus is prevented from inserting into the β-sheet in hGSi because the prosegment β-strand presumably remains fixed during the first two cleavages, preventing this re-arrangement until the final stages of conversion. Although the structure of mature gastricsin has not been determined, this assumption is supported by the position of the equivalent residues in mature pepsin. Furthermore, the position of the N-terminus is structurally conserved

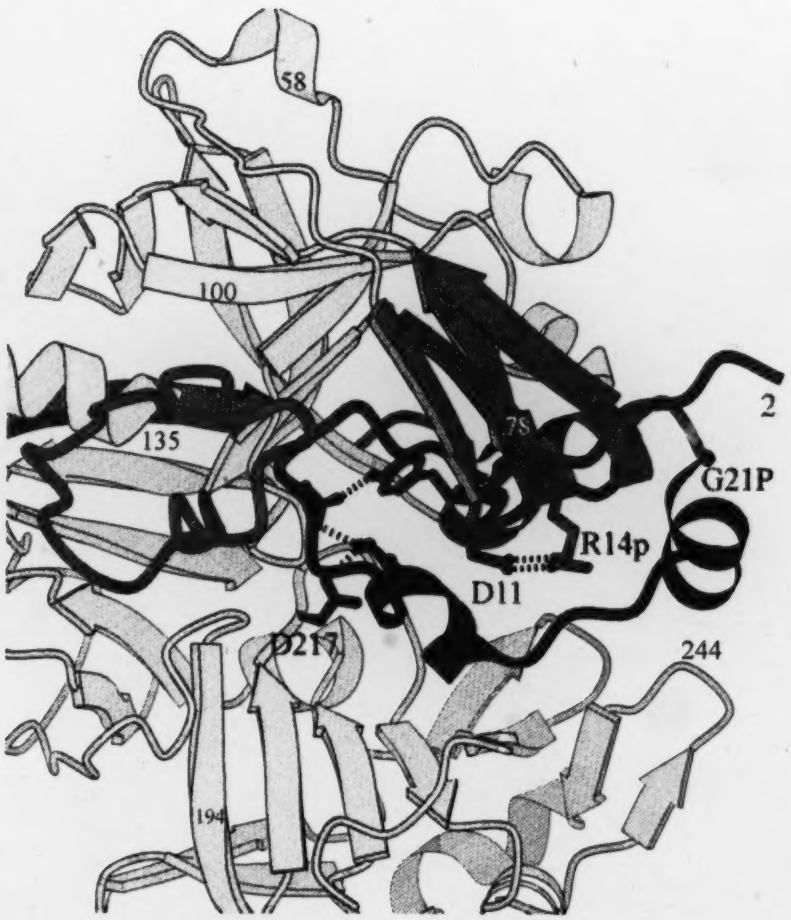


Figure 3C.3: Superposition of the zymogen (blue) and hGSi (red and grey), shown as ribbon models. The region of the prosegment of the zymogen that is cleaved and removed during the initial stages of activation is shaded in dark blue. The N-terminus of mature gastricsin in the zymogen and hGSi are separated by about 30 angstroms. Several loops in the vicinity of the active site, coloured in red (residues 2-13, 72-80 and 124-135) undergo major conformational re-arrangements during activation. The part of the prosegment that is coloured green (residues 1p-21p) is identical in the zymogen and hGSi.

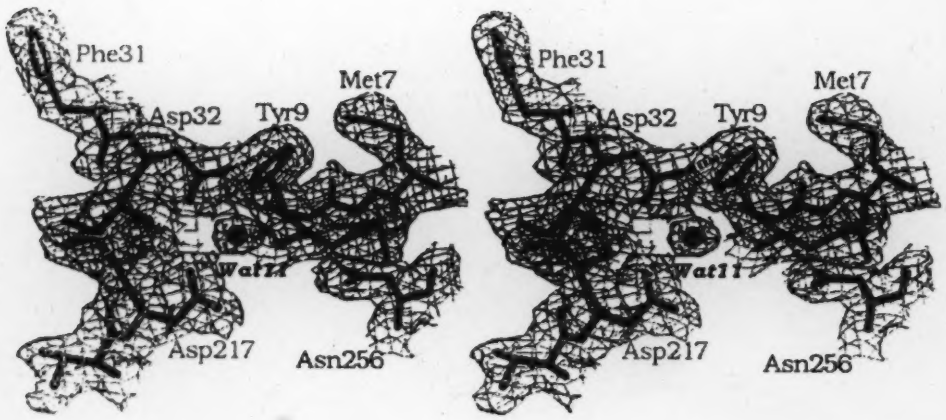


Figure 3C.4: Electron density at the active site of hGSi, shown in divergent stereo. The 2Fo-Fc map is contoured at 1.0s density. Hydrogen bonds are represented with dashes. The side chain of Asn256 is contributed by a symmetry-related molecule, and makes a hydrogen bond to the backbone nitrogen atom of Met10.

in the fungal enzymes (Yang et al., 1997; Cutfield et al., 1995), which share only about 25% sequence identity to pepsin and gastricsin. In summary, the remaining prosegment peptide (Ala1p-Phe26p) in hGSi indirectly obstructs the substrate binding cleft by preventing the insertion of the N-terminus of gastricsin into the β-sheet at the back of the enzyme.

Activation Mechanism. A mechanism is proposed for the acid activation of progastricsin that is consistent with the biochemical observations and with the structure of the intermediate. Exposure of the zymogen to low pH results in protonation of one of the Asp residues (Asp32 or Asp217) at the active site. The inhibitory salt bridges between Lys37p and the catalytic aspartate residues are thereby disrupted, leading to conformational changes that uncover the active site. Lys37p is located on a well-ordered 3_{10}-helix in the zymogen (Pro34p-Arg39p), but the flanking regions are highly mobile and have no interactions with mature gastricsin (Moore et al., 1995), suggesting that disruption of these several salt bridges will destabilize the entire region. In contrast to the 3_{10}-helix in hGSi, the prosegment β-strand (Val3p-Lys8p) is tethered to mature gastricsin by hydrogen bonds and hydrophobic interactions that are pH independent, and would be initially stable upon exposure to an acidic medium. In an insightful paper, Bent Foltmann predicted that the β-strand remains fixed during the formation of the analogous "intermediate 2" of porcine pepsin (Leu1p-Leu16p and mature pepsin; Foltmann et al., 1995). This hypothesis was supported by the inability to reconstruct the complex of the peptide with mature pepsin subsequent to the exchange of β-strands (Nielsen & Foltmann, 1993).

While the prosegment β-strand remains fixed in the zymogen, the three helices in the prosegment uncoil. A fourth helix at the N-terminus of the mature segment, spanning residues Glu5-Met10, is also disrupted. All of these helical regions are structurally conserved in porcine pepsinogen (James & Sielecki, 1986). This initial

conformational change results in the formation of 'intermediate 1' in the absence of a proteolytic cleavage. 'Intermediate 1' is detected by non-denaturing agarose gel electrophoresis upon brief acid activation of the zymogen on ice (Foltmann & Jensen, 1982). The conformational changes eventually result in the access of the Phe26p-Leu27p peptide bond to the active site (a movement of ~18 Å) and its cleavage by an intramolecular reaction. The newly formed N-terminus at Leu27p is flexible, and the second cleavage occurs at the pro-mature junction (Leu43p-Ser1), likely by an intermolecular mechanism (Foltmann & Jensen, 1982). The C-terminal prosegment peptide (Leu27p-Leu43p) is removed and hGSi is thus formed. The complex can be stabilized by re-forming the salt bridges between the prosegment peptide (Ala1p-Phe26p) and mature gastricsin upon elevation of the pH to neutrality. Perhaps the most important stabilizing salt bridge is that between Arg14p and Asp11 as it brings Tyr9(OH) to within hydrogen bonding distance (2.5 Å) of Asp32(Oδ2). In addition, the segment Ile13p-Lys20p has a strong propensity for a helical conformation, containing both N- and C-capping interactions (Richardson & Richardson, 1988; Harper & Rose, 1993), as well as a salt bridge between Lys11p and Glu19p along one face of the helix (Fig. 3C.2).

The intermediate complex regains activity below pH 5 (Foltmann & Jensen, 1982), although catalysis is inefficient because the substrate binding cleft has not fully formed. In order to understand how the intermediate complex becomes active, a model for the binding of the substrate-mimic pepstatin to the active site of hGSi has been constructed (Fig. 3C.8). Pepstatin is a potent inhibitor of pepsin and gastricsin (Fusek & Vetvicka, 1995; Davies, 1990), and during acid activation of pepsinogen, pepstatin is also able to bind activation intermediates and prevent the conversion to mature pepsin (Dykes & Kay, 1976; Glick et al., 1989). In the model of pepstatin-hGSi, a steric clash is observed between the phenolic ring of Tyr9 and the side chain of statine at the P1 position. Tyr9 is a conserved residue in all mammalian aspartic proteinases. Also, pepstatin residues at P3 (valine) and P4 (iso-valeryl) have steric

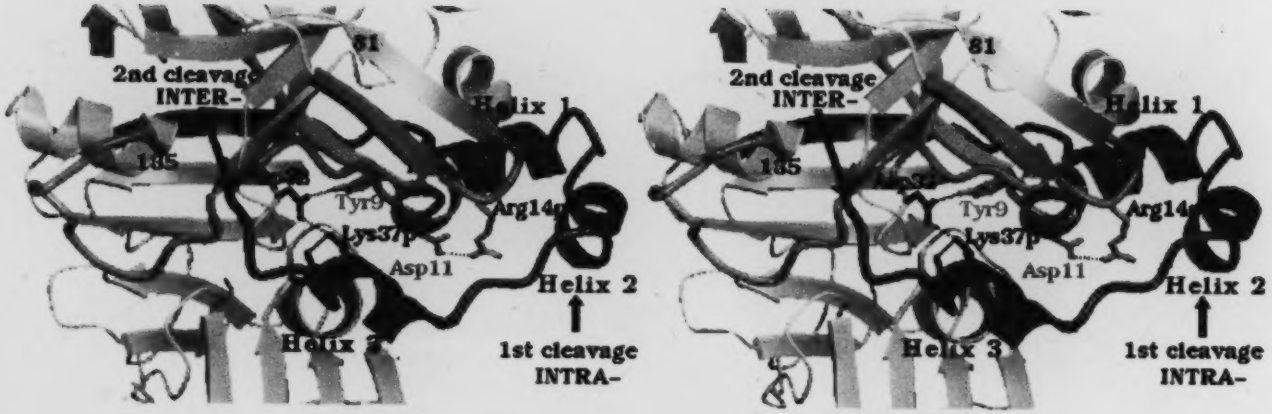


Figure 3C.5: Stereo ribbon model of the active site of progastricsin. The prosegment is green, and regions that undergo conformational changes are pink (residues 1-13, 72-80, and 124-135). The arrows indicate the sites of the first and second cleavage events that lead to the formation of hGSi, or 'intermediate 2'.

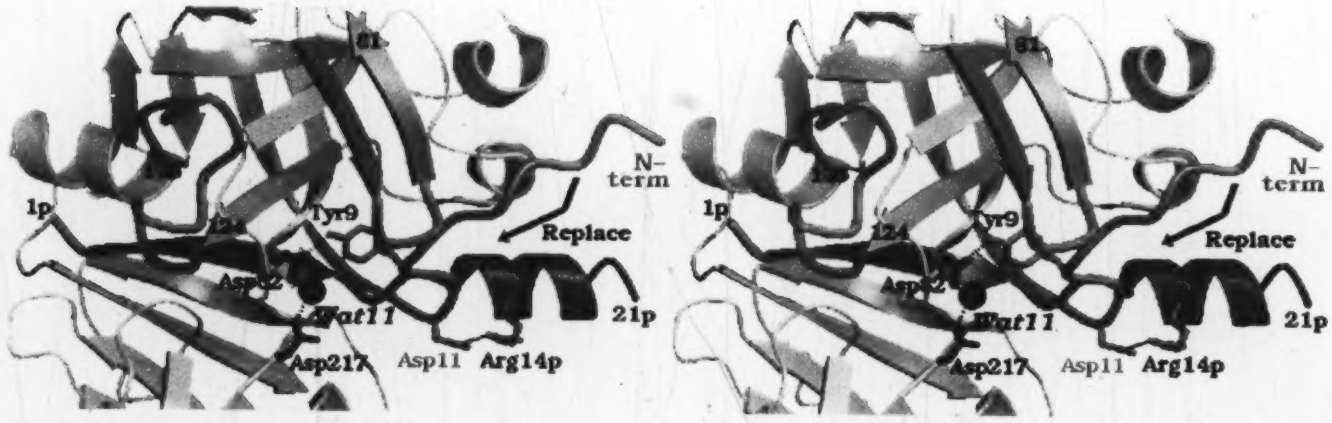


Figure 3C.6: Ribbon model of the active site of hGSi. The colouring scheme is similar to the preceding figures. The final stages of the activation pathway involve the replacement of the prosegment β-strand by the newly-formed N-terminus of mature gastricsin. The remaining prosegment of hGSi (green) presumably dissociates from the intermediate upon sufficiently long incubation times at low pH.

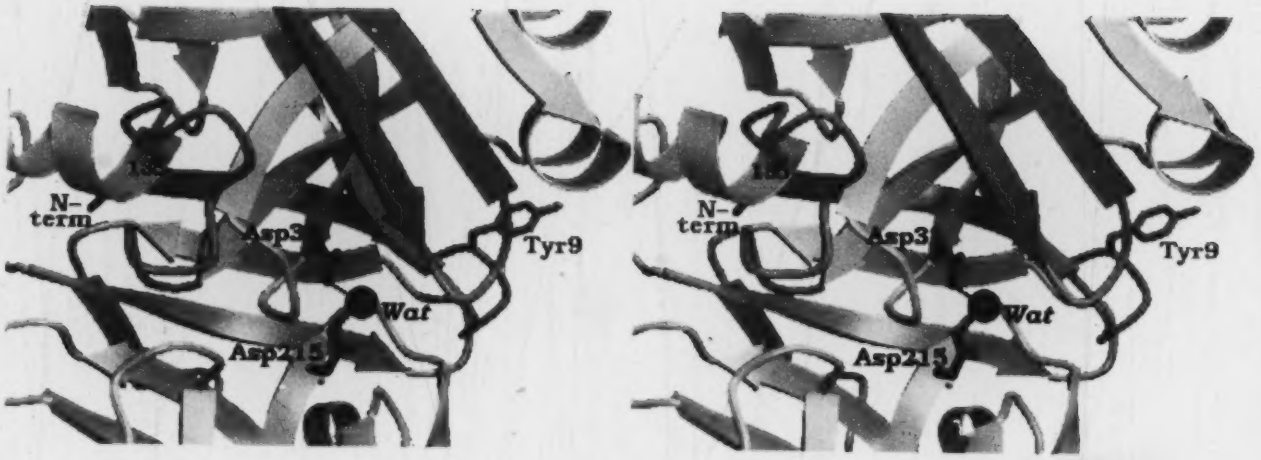


Figure 3C.7: Structure of active human pepsin, show in stereo as a ribbon model. The same colouring scheme as the previous figures is used. The prosegment has dissociated in the final stages of activation and is replaced by the mature N-terminus, which takes its position as the first strand in the six-stranded β-sheet.

clashes with Asp11 and Ala12. In contrast, residues at the P1'-P2' positions are accessible to the substrate binding cleft of hGSi. Upon lowering of the pH, it is evident from the model in Fig. 3C.8 that the segment Ala8-Ala12 must move in order to allow the substrates access to the S1-S3 binding pocket. The movement can be facilitated by disruption of salt bridges between Asp11-Arg14p, as well as Lys11p-Glu19p that would destabilize α1p. These conformational changes are likely reversible as long as the prosegment β-strand remains associated with the β-sheet of mature gastricsin. Prolonged exposure to an acidic environment favors the dissociation of the β-strand and its replacement by the N-terminus of mature gastricsin. The dissociation is presumably aided by the destablization of salt bridges between Arg14p-Asp11 (Fig. 3) and Lys4p-Asp172 (not shown), both of which tether the β-strand to the mature segment. The released prosegment peptide becomes rapidly degraded as mature gastricsin becomes fully active.

Comparisons to other zymogens. Activation intermediates have been observed in gastric aspartic proteinases from a variety of organisms that share sequence identities in the prosegments of the zymogens. However, the particular intermediate enzyme species and the kinetics of activation are diverse (Kageyama et al., 1989: Athauda et al., 1989; Marciniszyn et al., 1976). Given their sequence and structural similarities, it can be generalized that the prosegments of aspartic proteinases consist of acid-labile helical regions that rapidly vacate the active site allowing for transient proteolytic activity . They also have a more stable component, consisting of an N-terminal β-strand that ultimately dissociates from the β-sheet of the mature enzyme.

The variations in the cleavage sites and the kinetics of activation likely reflect the position and composition of prosegment regions that are accessible to intra- or intermolecular cleavage upon unfolding of the helices, as well as the relative strengths of the interactions between the prosegment β-strand and the mature enzyme. For example, the structure of the prosegment of pepsinogen brings the pro-mature

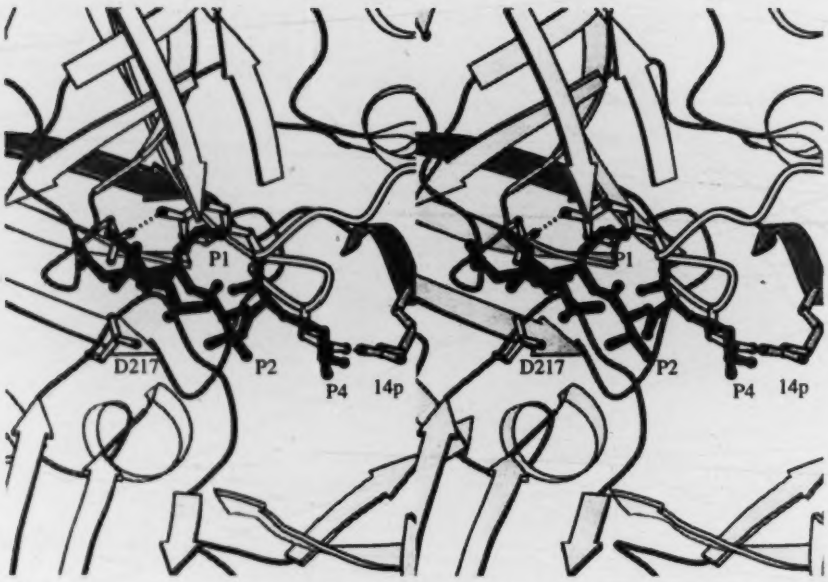


Figure 3C.8: Theoretical model of pepstatin at the active site of hGSi, in stereo. The model was constructed following optimal superposition of intermediate gastricsin with the pepsin-pepstatin complex (Fujinaga et al., 1995). The inhibitor is drawn with darkened bond vectors, and the prosegment ribbons of hGSi are also emphasized with dark shading. Clearly, the substrate-mimic pepstatin is unable to fit into the active site cleft due to steric clashes between the segment Tyr9-Asp11 and residues P1, P3 and P4 of the inhibitor.

junction to a solvent-exposed position on the surface of the molecule (Fig. 3C.9). Consequently, this junction is readily attacked by an active molecule of pepsin to generate mature pepsin and thereby release the full-length prosegment, which can be detected by various biochemical methods. In contrast, the prosegment of progastricsin follows a very different course, bringing the pro-mature junction closer to the surface of the zymogen. As a result, the pro-mature junction is not easily accessible to proteolysis *in trans*, which might explain why the full-length prosegment has never been detected and conversion appears to proceed exclusively *via* an intramolecular cleavage step.

Disruption of salt-bridge interactions is likely a common trigger for the activation of zymogens that are active in low pH compartments. For example, the structures of human protective protein (a serine carboxypeptidase) and procathepsins B/L (cysteine endopeptidases) suggest that the conversion process may be initiated by the disruption of conserved salt bridges (Rudenko et al., 1995; Coulombe et al., 1996; Cygler et al., 1996; Turk et al., 1996). In addition, intermediates have been detected during the trypsin-mediated activation of the human protective protein (Bonten et al., 1995) and yeast carboxypeptidase Y (Sorenson et al., 1994), implying that stepwise conversion of proteolytic precursors may be widespread. However, the forces that drive the conversion processes are not understood, and clearly further studies are required to appreciate the nature of the activation events from an energetic perspective.

Biological implications. It is not evident why self-activating gastric zymogens prefer the mechanism that has been described here. However, it can be reasoned that a stably-associated component of the prosegment (β1p), along with a gradual acquisition of catalytic activity, would be advantageous for preventing premature activation within the cell and during transport to the stomach. In addition, hydrolysis of the prosegment and its replacement by the N-terminus of mature gastricsin ensures

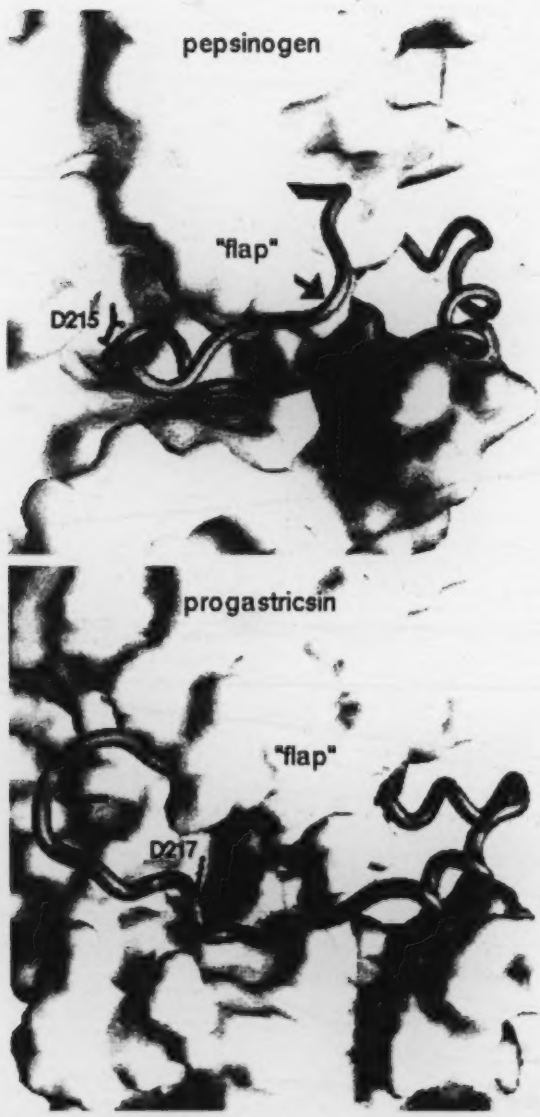


Figure 3C.9: Coil renderings of the prosegments of pepsinogen (top) and progastricsin (bottom) against their respective enzyme surfaces. This figure was drawn using the program GRASP (Nicholls et al., 1991). The arrow indicates the location of the pro-mature junction.

that the conversion is irreversible, and that the discarded prosegment peptides do not act as competitive inhibitors of the active enzyme. These properties are suited to the function of gastricsin, namely the non-specific degradation of acid-denatured dietary proteins in the stomach.

D. REFERENCES

Al-Janabi, J., Hartsuck, J.A. & Tang, J. (1972) Kinetics and mechanism of pepsinogen activation. *J Biol Chem* **247**: 4628-4632.

Athauda, S.B.P., Tanji, M., Kageyama, T. & Takahashi, K. (1989) A comparative study on the N-terminal amino acid sequences and some other properties of six isozymic forms of human pepsinogens and pepsins. *J Biochem* **106**: 920-927.

Bonten, E.J., Galjart, N.J., Wiilemson, N.J., Usmany, M., Vlak, J.M. & d'Azzo, A. (1995) Lysosomal protective protein/cathepsin A. Role of the "linker" domain in catalytic activation. *J Biol Chem* **270**: 26441-26445.

Brunger, A.T. (1993) "XPLOR: a system for X-ray crystallography and NMR", Yale University Press, New Haven, CT.

CCP4, in "The CCP4 Suite: Programs for Protein Crystallography". *Acta Cryst* **D50**: 760-763 (1994).

Coulombe, R., Grochulski, P., Sivarman, J., Menard, R., Mort, J.S. & Cygler, M. (1996) Structure of human procathepsin L reveals the molecular basis of inhibition by the prosegment. *EMBO J* **15**: 5492-5503.

Cutfield, S.M., Dodson, E.J., Anderson, B.F., Moody, P.C., Marshall, C.J., Sullivan, P.A. & Cutfield, J.F. (1995) The crystal structure of a major secreted aspartic proteinase from *Candida albicans* in complexes with two inhibitors. *Structure* **3**: 1261-1267.

Cygler, M., Sivarman, J., Grochulski, P., Coulombe, R., Storer, A.C. & Mort, J.S. (1996) Structure of rat procathepsin B. Model for inhibition of cysteine protease activity by the proregion. *Structure* **4**: 405-416.

Davies, D.R. (1990) The structure and function of the aspartic proteinases. *Annu Rev Biophys Biophys Chem* **19**: 189-215.

Dykes, C.W. & Kay, J. (1976) Conversion of pepsinogen into pepsin is not a one-step process. *Biochem J* **153**: 141-144.

Esnouf, R.M. (1997) An extensively modified version of MolScript that includes greatly enhanced colouring capabilities. *J Mol Graphics* **15**: 133-138.

Foltmann, B. & Jensen, A.L. (1982) Human progastricsin. Analysis of intermediates during activation into gastricsin and determination of the amino acid sequence of the propart. *Eur J Biochem* **128**: 63-70.

Foltmann, B. Aspartic Proteinases: alignment of amino acid sequences, in *Proceedings of the 18th Linderstrom-Lang Conference, 4-8 July 1988, Elsinore, Denmark* (ed. B. Foltmann) 7-20 (1988).

Foltmann, B., Harlow, K., Houen, G., Nielsen, P.K. & Sangild, P. Comparative investigations on pig gastric proteinases and their zymogens, in *Aspartic Proteinases: Structure, Function, Biology and Biomedical Implications* (ed. K. Takahashi) 41-51 (Plenum Press, New York, 1995).

Fujinaga, M., Chernaia, M.M., Tarasova, N.I., Mosimann, S.C. & James, M.N.G. Crystal structure of human pepsin and its complex with pepstatin. (1995) *Protein Sci* **4**: 960-972.

Fusek, M. & Vetvicka, V. Aspartic Proteinases: Physiology and Pathology, (CRC Press, New York, New York, 1995).

Glick, D.M., Shalitin, Y. & Hilt, C.R. (1989) Studies on the irreversible step of pepsinogen activation. *Biochemistry* **28**: 2626-2630.

Harper, E.T. & Rose, G.D. (1993) Helix stop signals in proteins and peptides: the capping box. *Biochemistry* **32**: 7605-7609.

James, M.N.G. & Sielecki, A.R. (1986) Molecular structure of an aspartic proteinase zymogen, porcine pepsinogen, at 1.8Å resolution. *Nature (London)* **319**: 33-38.

Jones, T.A. & Kjeldgaard, M. (1995) "O - the Manual", Version 5.11, Uppsala, Sweden.

Kageyama, T. & Takahashi, K. (1984) Rabbit pepsinogens. Purification, characterization, analysis of the conversion process and determination of the N-terminal amino-acid sequences. *Eur J Biochem* **141**: 261-269.

Kageyama, T., Ichinose, M., Miki, K., Athauda, S.B., Tanji, M. & Takahashi, K. (1989) Differences of activation processes and structure of activation peptides in human pepsinogens A and progastricsin. *J Biochem* **105**: 15-22.

Marciniszyn, Jr., J., Huang, J.S., Hartsuck, J.A. & Tang, J. (1976) Mechanism of intramolecular activation of pepsinogen. *J Biol Chem* **251**: 7095-7102.

Merritt, E.A. & Murphy, M.E.P. (1994) Raster3D version 2.0: a program for photorealistic molecular graphics. *Acta Cryst* **D50**: 869-873.

Moore, S.A., Sielecki, A.R., Chernaia, M.M., Tarasova, N.I. & James, M.N.G. (1995) Crystal and molecular structures of human progastricsin at 1.62Å resolution. *J Mol Biol* **247**: 466-485.

Navaza, J. (1994) AMoRe: an automated package for molecular replacement. *Acta Cryst* **A50**: 157-163.

Nicholls, A., Sharp, K.A. & Honing, B. (1991) Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins* **11**: 281-296.

Nielsen, F.S. & Foltmann, B. (1993) Activation of porcine pepsinogen A: the stability of two non-covalent intermediates at pH 8.5. *Eur J Biochem* **217**: 137-142.

Otwinowski, Z. in *Data Collection and Processing* (eds Sawyer, L., Isaacs, N. & Bailey, S.) (SERC Daresbury Laboratory, Warrington, UK, 1993).

Pannu, N.S. & Read, R.J. (1996) Improved structure refinement through maximum likelihood. *Acta Cryst* **A52**: 659-668.

Richardson, J.S. & Richardson, D.C. (1988) Amino acid preferences for specific locations at the ends of α-helices. *Science* **240**: 1648-1652.

Rudenko, G., Bonten, E., d'Azzo, A. & Hol, W.G.J. (1995) Three-dimensional structure of the 'human protective protein': structure of the precursor form suggests a complex activation mechanism. *Structure* **15**: 1249-1259.

Sogawa, K., Fuji-Kuriyama, Y., Mizukami, Y., Ichihara, Y. & Takahashi, K. Primary structure of human pepsinogen gene. *J. Biol. Chem.* **258**, 5306-5311 (1983).

Sorenson, S.O., Van den Hazel, H.B., Kielland-Brandt, M.C & Winther, J.R. (1994) pH-dependent processing of yeast carboxypeptidase Y by proteinase A *in vivo* and *in vitro*. *Eur J Biochem* **220**: 19-27.

Turk, D., Podobnik, M., Kuhelj, R., Dolnar, M. & Turk, V. (1996) Crystal structures of procathepsin B at 3.2 and 3.3Å resolution reveal an interaction motif between a papain-like cysteine protease and its propeptide. *FEBS Lett* **384**: 211-214.

Yang, J., Teplyakov, A. & Quail, J.W. (1997) Crystal structure of the aspartic protease from *Rhizomucor miehei* at 2.15Å resolution. *J Mol Biol* **268**: 449-459.

CHAPTER 4

GENERAL DISCUSSION AND CONCLUSIONS

Historical perspective. The study of pepsin is a dramatic and fascinating story in the history of medicine. Perhaps the earliest written account of the protein comes from the work of William Beaumont, a surgeon in the United States army. In the early 1830s, Beaumont's fortunate encounter with a gunshot victim, a Canadian Metis named Alexis St. Martin, led to a pioneering case study of the digestive process. The patient developed a gastric fistula from imperfect closure of the wound, allowing Beaumont to perform chemical analyses of the gastric juice during digestion. In 1833, Beaumont published his "Experiments and Observations on Gastric Juice and the Physiology of Digestion" (Myer, 1912) in which he confirmed the presence of hydrochloric acid in the stomach, but also identified an "unknown chemical substance". It was a few years later that Schwann used the term pepsin for this chemical substance, demonstrating its power to change insoluble albumins into soluble "peptones" (from the Greek *peptos*, meaning digested or cooked).

In modern day laboratories, techniques in biochemistry, molecular biology, and structural biology are now routine tools for addressing problems in the field of protein structure and function. The importance of these enzymes in diverse biological processes, and especially their relevance to human diseases, ensures that the story of the aspartic proteinases will continue to unfold. The discussion below is an elaboration of the work described in this thesis and its implications for some of the fundamental problems with regard to the aspartic proteinases, as well as protein structure and function in general. Finally, some fascinating and recent developments are highlighted that point to future directions in the aspartic proteinase field.

A. Ultra-high resolution crystal structure of PP7-penicillopepsin.

The structure of PP7-penicillopepsin is the first "atomic" resolution model for an aspartic proteinase. Atomic resolution is defined by George Sheldrick (1990) as data extending to 1.2 Å resolution with 50% of the intensities >2.0σ in the highest resolution shell. The refinement was performed using data to 0.90 Å, which is significantly shorter than the chemical bonds between atoms. However, the data are relatively incomplete in the 0.95-0.90 Å bin, so that the true resolution of the structure is considered to be 0.95 Å resolution. Aside from the penicillopepsin structures from our lab (Ding et al., 1998), the highest reported resolution that has been attained for an aspartic proteinase is the 1.60 Å structure of endothiapepsin in complex with various inhibitors (Bailey & Cooper, 1994). The structures of active penicillopepsin and PP6-penicillopepsin are also being refined beyond 1.0 Å resolution, and it will be interesting to compare the active and inhibited forms of the enzyme in the future (Marie Fraser, unpublished data).

The 2973 non-hydrogen atoms in the model of PP7-penicillopepsin represent one of the largest molecules for which an atomic structure has been determined. Consequently, the structure provides a wealth of information to enhance our general understanding of protein architecture. As of August, 1998, the Brookhaven Protein Data Bank contained 5731 crystallographic entries with coordinate lists of proteins larger than 100 residues. A search of this subset of proteins revealed that there are only 12 structures that have been determined using data better than 1.20 Å resolution (Table 4A.1). Only the phosphate-binding protein (entry 1IXH) can be considered on a par with PP7-penicillopepsin in regard to size, resolution, and the quality of refinement.

The 0.98 Å structure of phosphate-binding protein (PBP) represents a useful comparison to PP7-penicillopepsin. The complex of phosphate with PBP was crystallized at pH 4.5, the same pH at which crystals of PP7-penicillopepsin were

Table 4A.1: High-Resolution Crystal Structures of Proteins

protein (PDB code)	No. of atoms protein/water	resolution	R_{cryst}/R_{free} (program)	reference
cutinase (1CEX)	1440/264	1.0 Å	9.4/11.9 (SHELX-97)	Longhi et al., 1997
γ-β-crystallin ((1AMM)	1486/392	1.20 Å	18.5/--- (RESTRAIN)	unpublished
protease I (1ARB)		1.20 Å	14.9 PROLSQ	Tsunasawa et al., 1989
spectrin β-chain (1BKR)	887/208	1.10 Å	14.1/18.7 SHELX-97	unpublished
subtilisin-eglin C (1CSE)		1.20 Å	17.8/--- "ERE"	Bode et al., 1987
fatty-acid binding protein (1IFC)	1057/238	1.19 Å	16.9/--- X-PLOR,TNT	Scapin et al., 1992
phosphate binding protein (1IXH)	2448/703	0.98 Å	11.4/14.1 SHELX-96	Wang et al., 1997
concanavalin A (1NLS)	1809/323	0.94 Å	12.7/15.4 SHELX-97	Deacon et al., 1997
oligopeptide binding prot. (1JET)	4193/560	1.20 Å	22.9/25.5 (PROLSQ)	Tame et al., 1996
Tyr. kinase SH2 domain (1LKK)		1.0 Å	13.3/--- (SHELX-93)	unpublished
peptidyl t-RNA hydrolase (2PTH)	1523/184	1.20 Å	19.6/21.5 X-PLOR	Schmitt et al., 1997
lysozyme (1JSE)	1016/157	1.12 Å	10.40/14.0 (SHELX-93)	Harata et al., 1998

grown. The high-resolution structure of phosphate-PBP provided the first definitive proof of a short hydrogen bond in a protein-ligand complex. A distance of 2.432 Å is observed between the Oδ2 atom of Asp56 and the oxygen atom of inorganic phosphate (Fig. 4A.1). The positions of these atoms are accurate to less than 0.01 Å (Wang et al., 1997). However, it is interesting to note that the position of the hydrogen atom that is shared by the Oδ2 atom of Asp56 and the phosphate oxygen was not identified. It was not apparent from the difference electron density maps in the paper (Wang et al., 1997), so that it is unclear whether the hydrogen resides on the phosphate (i.e., P-O-H----Oδ2) or the carboxylate of Asp56 (i.e., P-O----H-Oδ2).

The high resolution structure of PP7-penicillopepsin represents the second protein-ligand complex that unambiguously demostrates the presence of a short hydrogen bond. An initial unrestrained full matrix refinement of the structure indicates that the positions of the non-hydrogen atoms at the active site are accurately determined, with atomic positional estimated-standard-deviations (esds) less than 0.01 Å. Previous crystallographic studies of penicillopepsin-inhibitor complexes have involved extensive analyses of the non-bonded contacts between oxygen atoms at the active site in order to determine possible hydrogen atom positions (Fraser et al., 1992). From these studies, a hydrogen is placed between the *pro-S* oxygen atom of the phosphonate and the Oδ1 atom of Asp33. Also, the Oδ1 atom of Asp213 is a hydrogen bond donor to the *pro-R* oxygen of the phosphonate. The refined distances corresponding to the bond lengths and non-bonded contacts at the active site of PP7-penicillopepsin are consistent with this hydrogen bond network (Fig. 4A.2). The bond distances shown in this figure were not retrained during the latter stages of refinement in SHELX-97. The C-O bond lengths of the carboxylate groups of Asp33 and Asp213 converge to single bonds (1.304 Å) or double bonds (1.231 Å), according to the parameters defined by Engh & Huber (1991). It is disappointing that the hydrogen atoms presumed to mediate the interactions between these oxygens are

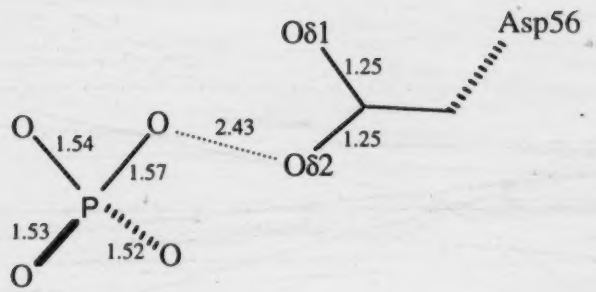


Fig. 4A.1: A schematic of the short hydrogen bond between the phosphate-binding protein and phosphate. The PDB code for the corodinates is 1IXH (Wang et al., 1997). The hydrogen bond is formed between Asp56 of the protein and the phosphate oxygen. The hydrogen atoms were not addressed in the paper.

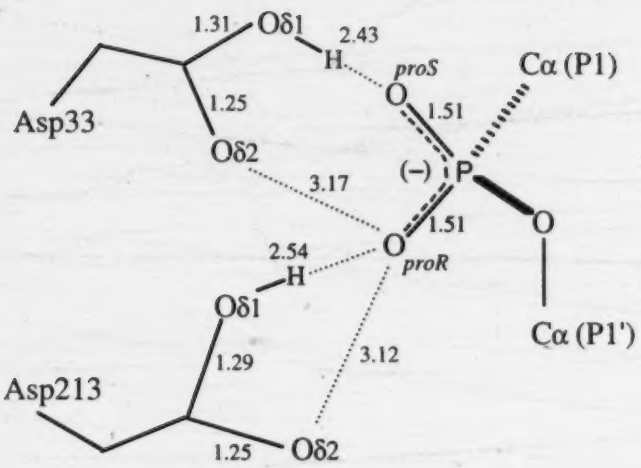


Fig. 4A.2: A schematic of the hydrogen-bonding scenario at the active site of PP7-penicillopepsin. The hydrogen atom positions are assigned as previously described (Fraser et al., 1992). The bond distances are shown, and dotted lines denote non-bonding contacts between oxygen atoms.

not observed directly in the electron density maps of PP7-penicillopepsin (see below). However, hydrogen atoms are often difficult to assign with confidence even in small molecule crystal structures.

The importance of short hydrogen bonds in biological interactions remains a matter of controversy. In the serine proteinases, a short hydrogen bond between the backbone NH atom of the P1 residue and the backbone carbonyl oxygen of Ser214 was believed to enhance the stabilization of the transition state, as compared to a normal hydrogen bond (Robertus et al., 1972). However, mutational analysis of the phosphate-binding protein revealed limited effects of the short hydrogen bond upon the binding affinity to phosphate. The aspartate residue (Asp56) involved in the short contact with phosphate was mutated to Asn, and the crystal structure of the mutant PBP with phosphate was determined (Wang et al., 1997). The structure revealed a conventional 2.6 Å hydrogen bond between the Oδ2 atom of Asn56 and the phosphate oxygen, but the binding affinity was virtually identical to the wild-type protein.

Hydrogen atoms. The positions of many hydrogen atoms are apparent from difference electron density maps of PP7-penicillopepsin. The F_o-F_c 'omit map' shown in Fig. 4A.3 was calculated following removal of all hydrogens and 10 cycles of conjugate-gradient least -squares minimization, after which the R-factor increased by 1.4%. The most visible hydrogens are those bound to the C_α carbons and the backbone nitrogens of residues that form stable β-sheets. The positions of potential hydrogen atoms at the active site are ambiguous in the difference maps. For example, a consistent peak is observed between Asp33(Oγ1) and the *pro-S* oxygen atom of the phosphonyl group (Fig. 4A.4). However, the distance and geometry of this peak cannot be reconciled with the symmetric hydrogen bond that is expected to lie midway between these oxygen atoms.

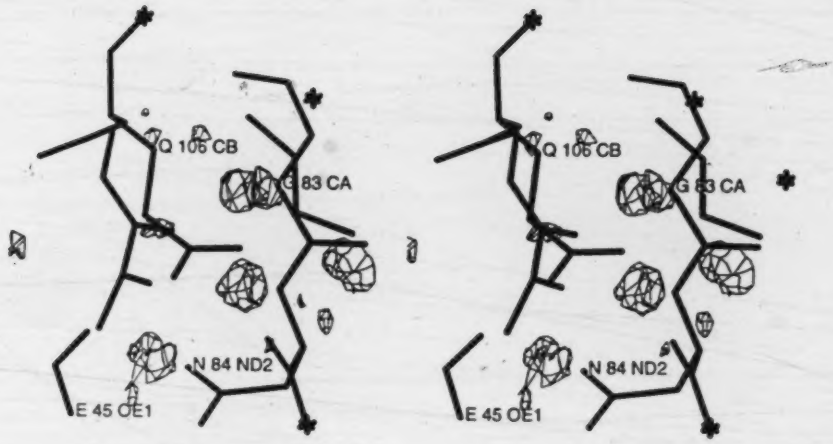


Figure 4A.3: A section of the difference electron density 'omit' map of PP7-penicillopepsin. The map is was calculated following removal of hydrogen atoms and a further 10 cycles of conjugate gradient least-squares minimization using SHELX-97. The Fo-Fc map is contoured at 2.5σ. Density corresponding to several hydrogens are visible, including the Cα hydrogens of Gly83, the backbone NH hydrogen of Asn84, and the sidechain hydrogen of the Nδ2 atom which is hydrogen bonded to the side chain of Glu45(Oε2).

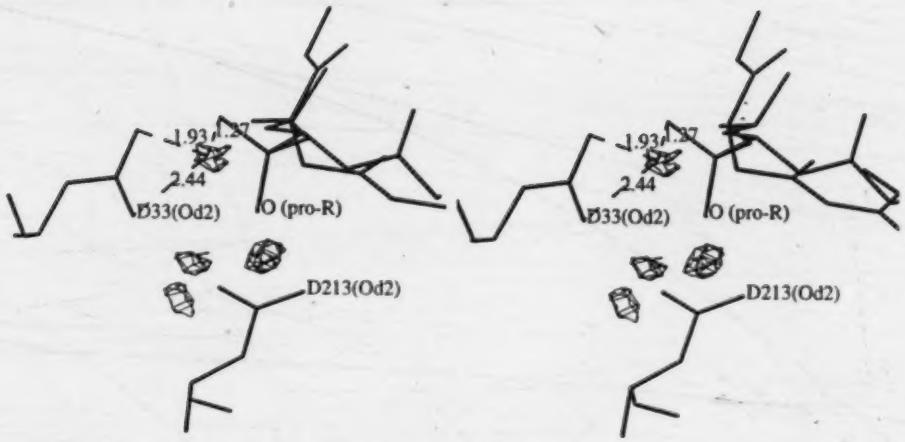


Figure 4A.4: A section of the difference electron density 'omit' map of PP7-penicillopepsin at the active site. Dashed lines indicate the distances to the electron density peak between Asp33 and the phosphonate oxygen. This Fo-Fc map was calculated with hydrogen atoms present during Fourier calculations. However, hydrogens were excluded from the side chains of aspartate and glutamate residues. The map is contoured at 2.0σ.

Deviations from planarity of peptide bonds. In their classical paper, Corey and Pauling (1953) proposed the fundamental concept of the *trans* configuration and planar nature of peptide bonds in proteins. However, deviations from planarity of peptide bonds have since been observed both in small linear peptides and in atomic models of proteins (Naganathan & Venkatesan, 1972; MacArthur & Thornton, 1996). Indeed, it was realized very early that deviations from planarity are necesary to accomodate the conformational space that is occupied by a polypeptide (Ramachandran, 1968).

The atomic structure of PP7-penicillopepsin allows accurate determination of the deviations from planarity of peptide bonds. At low resolution, the least-squares refinement process is dependent upon geometric restraints that are imposed in order to supplement the diffraction data, since the observation-to-parameter ratios are typically low. For example, the 2.3 Å structure of the gastricsin intermediate (Chapter 3) contained 5207 atoms in the asymmteric unit, and this model was refined against 34,124 reflections. Therefore the observation:parameter ratio was 34,124/(4 x 5207) = 1.64. The multiplication by a factor of 4 comes from the (x,y,z) co-ordinates and the B-factors that are refined for each atom. The model of PP7-penicillopepsin was refined with anisotropic *B*-factors, but even allowing for these additional parameters, the observation-to-parameter ratio was 5.6.

Many of the peptide bonds in the refined model of PP7-penicillopepsin are significantly distorted from planarity. Table 4A.2 lists the mean and standard deviations of the peptide bond geometry among various penicillopepsin models. The ω-angle is the torsion angle that is formed by four consecutive atoms, $C_{\alpha,i}$-C_i-N_{i+1}-$C_{\alpha,i+1}$, along a polypeptide chain. The peptide bond linking residues Gln306 and Tyr307 provides a concrete example of severe distortion from planarity (Table 4A.2, Fig. 4A.5). This peptide bond is the most distorted of any found in PP7-penicillopepsin, but the structure is well ordered within this region. In fact, the

Table 4A.2: Statistics of the ω-angle and its Distortion from Planarity in Penicillopepsin Structures

penicillopepsin complex	refinement program†	<ω> (°)	standard deviation (°)	306-307 ω-angle
*Native	PROLSQ	179.5	3.0	169
PP2	X-PLOR/ PROLSQ	179.5	3.9	167
PP7	X-PLOR/ SHELX-97	179.2	7.0	154
PP8	X-PLOR/ SHELX-97	179.4	6.2	155

The statistics were derived from the peptide bonds that link the 323 amino acids comprising the protein. Inhibitor atoms were excluded from the calculations. The "306-307" ω-angle corresponds to the peptide bond linking residues Gln306-Tyr307.

*The structure of native penicillopepsin, PDB code 3APP (James & Sielecki, 1983).

†The references for the programs are: PROLSQ, Hendrickson & Konnert (1980); X-PLOR, Brunger (1992); and SHELX-97, Sheldrick (1997).

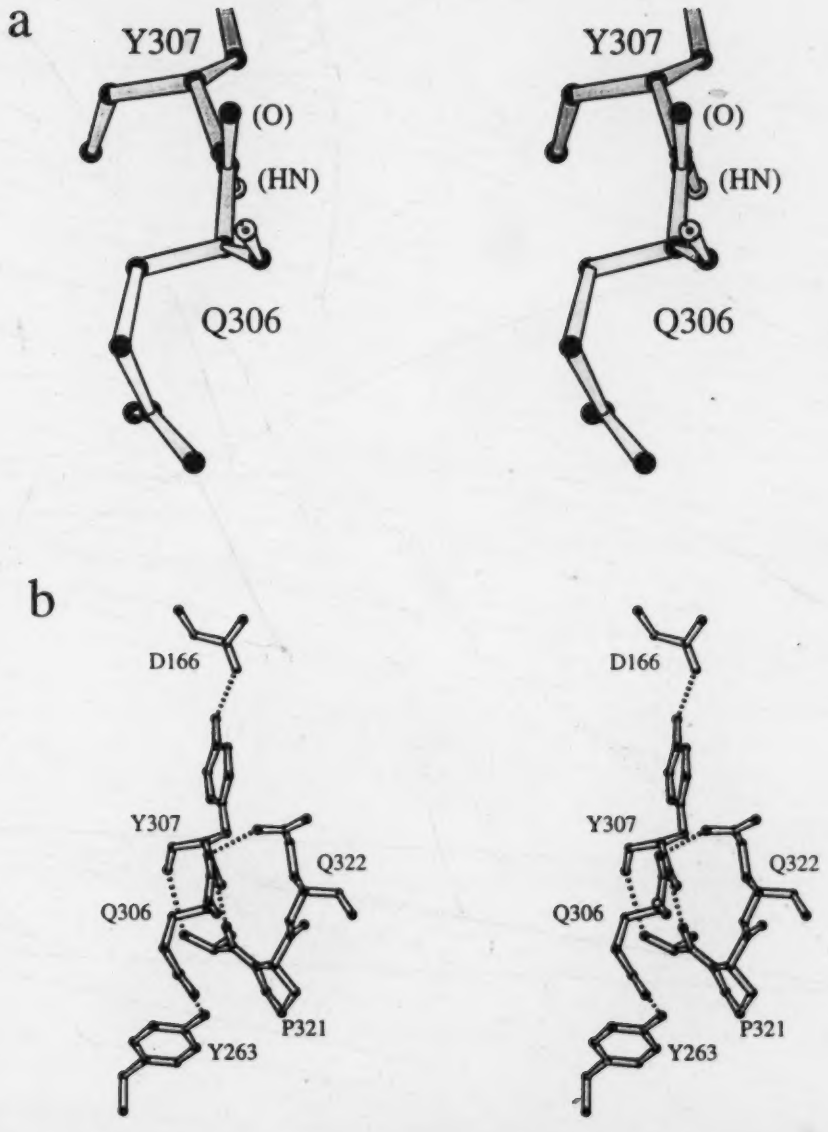


Figure 4A.5: Ball-and-stick model of PP7-penicillopepsin at the Gln306-Tyr307 peptide bond. (a) Close-up of the ω-angle [Cα(306)-N-C-Cα(307)], which is significantly distorted from an ideal plane. The carbonyl oxygen and the amide proton are labelled. (b) The molecular environment of the peptide bond. Hydrogen bonds and hydrophobic interactions adjacent to the peptide bond likely compensate for the unfavourable torsion angle. Dashed lines indicate hydrogen bonds.

deviation of this peptide bond from planarity was apparent from the very first difference density maps that were calculated during the initial stages of refinement in X-PLOR.

The residues Gln306 and Tyr307 are found at the termination of a short helix and the beginning of a β-strand (residues 307-311) that lies in the middle of the 6-stranded β-sheet behind the active site. The backbone oxygen and nitrogen atoms in this peptide bond make several hydrogen bonds with neighbouring atoms (Fig. 4A.3). One of the most ordered water molecules in the structure, WAT3, is part of a network of hydrogen bonds that are observed in this region. Tyr307 is involved in hydrophobic interactions, in part with Tyr175, and the side chain hydroxyl atom of Tyr307 makes a 3.0 Å hydrogen bond with the backbone NH atom of Asp171. It is likely that the favorable hydrogen-bonding and hydrophobic interactions in the vicinity of Gln306 and Tyr307 compensate for the distorted peptide bond that is observed.

Conventional programs such as X-PLOR typically apply tight planarity restraints upon peptide bonds during refinement. However, the small number of proteins that have been determined at atomic resolution reveal that significant deviations of peptide bonds from ideal planes are common (Dauter et al., 1997). These observations are consistent with the distribution of ω-angles for peptides found in the Cambridge Structural Database (Morris et al., 1992), suggesting that protein refinement programs need to soften the planarity restraints to reflect these distortions of peptide bonds. In this respect, the high-resolution structures of penicillopepsin and other proteins are an important resource for obtaining better estimations of chemical and stereochemical parameters that can be implemented in refinement programs.

The reasons underlying severe distortions of the ω-angle away from 180° in proteins is not well understood. Model calculations indicate that secondary structure

elements which involve hydrogen bonds with near-ideal geometry (such as an α-helix) result in peptide bonds that resist distortions from planarity (Scheiner & Kern, 1977). In contrast, β-strands and other conformations contain less regular hydrogen-bonding geometry (MacArthur & Thornton, 1996), which may promote distortions of the peptide bond. The structural context of the distortions from planarity of peptide bonds in the penicillopepsin structures determined in our lab are currently being studied in detail.

The PP7-penicillopepsin complex as a transition-state mimic. The transition-state mimicry exhibited by the phosphonate inhibitors and fluorine-substituted pepstatin inhibitors in complexes with penicillopepsin has been discussed in detail previously (James et al., 1992; Fraser et al., 1992). Both groups of inhibitors position a mimic of the tetrahedral transition-state intermediate at the active site. However, the P1-statine moieties of the pepstatin-like inhibitors contain additional main-chain atoms (see Fig. 1B.5). In this respect, the phosphonate inhibitors are more faithful representatives of the P1-P1' positions of good substrates (Fraser et al., 1992). In addition, the torsion angle of the bound phosphonate inhibitors that is equivalent to the ω-angle of the P1 residue in a normal substrate resembles more closely the antiperiplanar conformation (~180°) which is expected in the transition state (James et al., 1992). In the PP7-penicillopepsin complex, this ω-angle is 147° (see Table 2C.4, p. 68), whereas the corresponding ω-angle is 70°-80° in the pepstatin-like inhibitors (James et al., 1992; Fraser et al., 1992).

Analysis of disorder in PP7-penicillopepsin. The high-resolution structure of PP7-penicillopepsin allow a detailed analysis of the motion of the protein. Fig4A.6 is a bar graph representation of several motion-related parameters derived from the final unrestrained least-squares cycle, plotted against the residue numbers. Visual

inpection of this figure permits a qualitative analysis of the anisotropic motion and disorder. It is apparent that the loop comprising residues 277-282 is severely disordered. This region was the only part of the model for which the backbone atom positions could not be determined reliably.

The bar graph in panel 'b' of the figure depicts the anisotropic motion of the main chain (averaged per residue), with a value of 1.0 representing completely isotropic motion. Interestingly, the 'flap' region of penicillopepsin (residues ~63-80) exhibits among the highest anisotropic motion in the structure. The flexibility of this equivalent flap region in 'intermediate 2' of human gastricsin was dramatically revealed upon comparison of this structure to the zymogen, progastricsin (Chapter 2). As discussed previously, the mobility of the flap is also apparent from comparisons of the structures of native aspartic proteinases with those containing bound inhibitors. The flap closes like a lid, clamping the inhibitor to the active site. The anisotropic character of the flap in the structure of PP7-penicillopepsin may reflect the propensity of this region to undergo this opening and closing motion. Interestingly, anisotropic motion in the high resolution crystal structure of the enzyme cutinase suggests a similar role in substrate binding (Longhi et al., 1997), consistent with parallel NMR studies of cutinase (Pompers et al., 1996).

It is also worthy to note that a helix adjacent to the flap (residues 110-115; blue region in panel 'b') similarly displays anisotropic motion, relative to the remainder of the protein. This helix connects two β-strands that are othogonal to each other and form part of the β-barrel in the N-terminal domain. Many of the serine side chains are surface-exposed in the enzyme and are disordered (grey lines; panel 'd' in Fig. 4A.6). In summary, the quality of the PP7-penicillopepsin model is excellent as evidenced by the low esds that were derived from unrestrained full-matrix refinement (panel 'c').

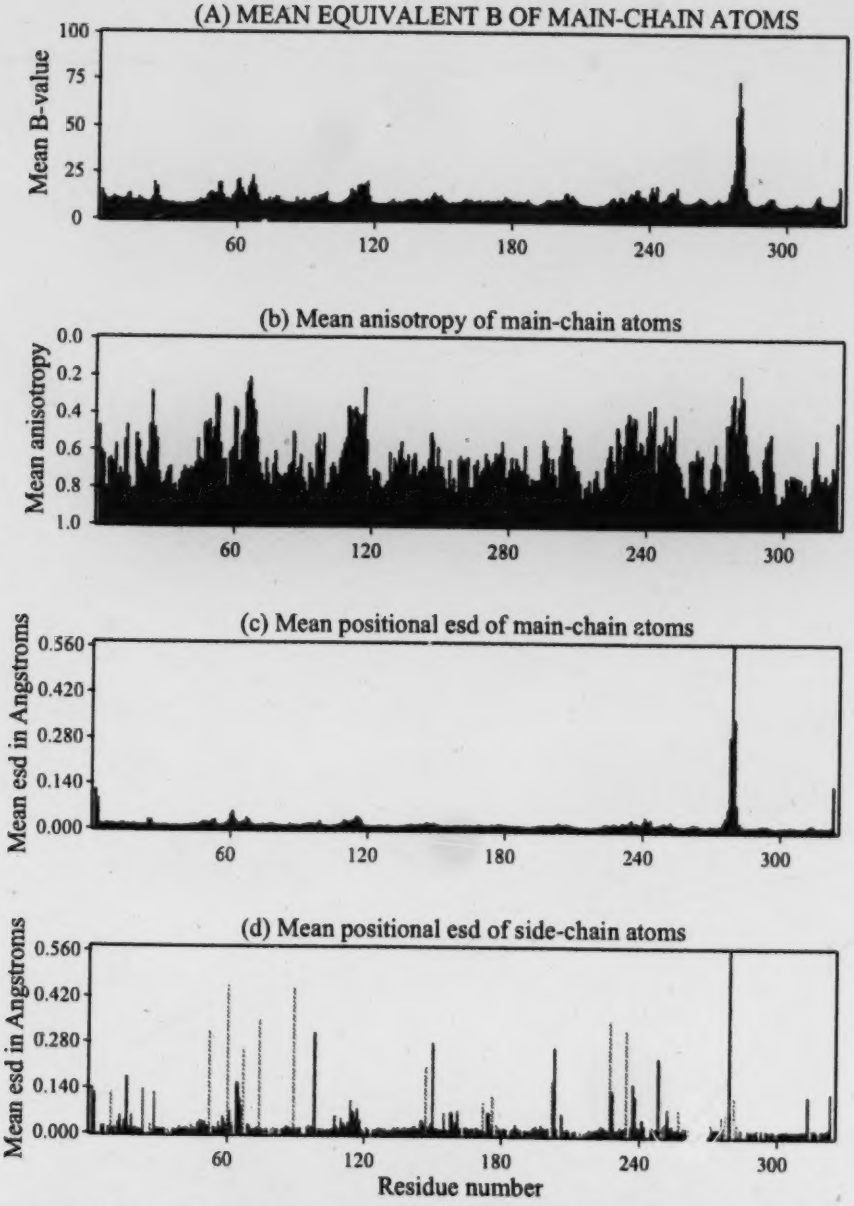


Figure 4A.6: Analysis of disorder in PP7-penicillopepsin. The residue numbers are plotted against (a) mean equivalent B-value of main chain atoms, (b) mean anisotropy of main chain atoms, (c) mean positional estimated standard deviations (esds) of main chain atoms, and (d) mean positional esds of side chain atoms. In plots a-c, green corresponds to β-strands, blue regions are α-helices, and red denotes residues in other secondary structures. In plot (d), green = Phe, Tyr, Trp, His; yellow = Cys, Met; cyan = Gly, Ala, Leu, Ile, Val, Pro; purple = Gln, Asn; grey = Ser, Thr; blue = Arg, Lys; and red = Glu, Asp.

B. Conversion of zymogens to active proteolytic enzymes.

The inhibitory interactions utilized by the prosegments of aspartic proteinases to prevent proteolytic activity are distinct from the manner in which peptide-based competitive inhibitors interact with the active enzymes. In both cases, the active site cleft of the enzyme is preformed and competent for catalyzing the cleavage of peptide bonds. However, the prosegment interacts with the active site via a 3_{10}-helix (α3p), whereas the phosphonate inhibitors stretch across the active site cleft in a β-strand conformation. The prosegment-active site interactions involve a few hydrogen bonds and one key salt bridge (Lys37p in hPGC). In contrast, the phosphonate inhibitors form extensive hydrogen bonding and hydrophobic interactions that bury more than 80% of the inhibitor.

These properties of prosegments vs. transition-state inhibitors can be generalized to other proteolytic enzymes. The inhibitory interactions between the prosegment and the active site are generally loose and poised for disruption (Fig. 4B.1), unlike small molecule or peptide-based transition-state mimics which bind tightly and often mimic the cleavage transition state. This makes sense from a biological perspective because the prosegments must be removed during conversion, following the appropriate stimuli (*e.g.*, pH changes, limited proteolysis).

The differences in the mechanism of inhibition can also be extended to prosegments *vs.* the so-called "protein-proteinase" inhibitors of proteolytic enzymes. Following the conversion process, there exists an elaborate collection of specialized proteins that specifically inhibit the active enzymes. The importance of these 'protein-proteinase' inhibitors range from attenuation of the serine proteinase cascades (blood coagulation, complement activation) to a role in the defense mechanism against foreign pathogens. For the serine proteinases, the two families of these inhibitors are the substrate-like 'canonical' inhibitors (Bode & Huber, 1992; Laskowski & Kato, 1980) and the 'serine proteinase inhibitors' (serpins; Wright, 1996; Potempa et al.,

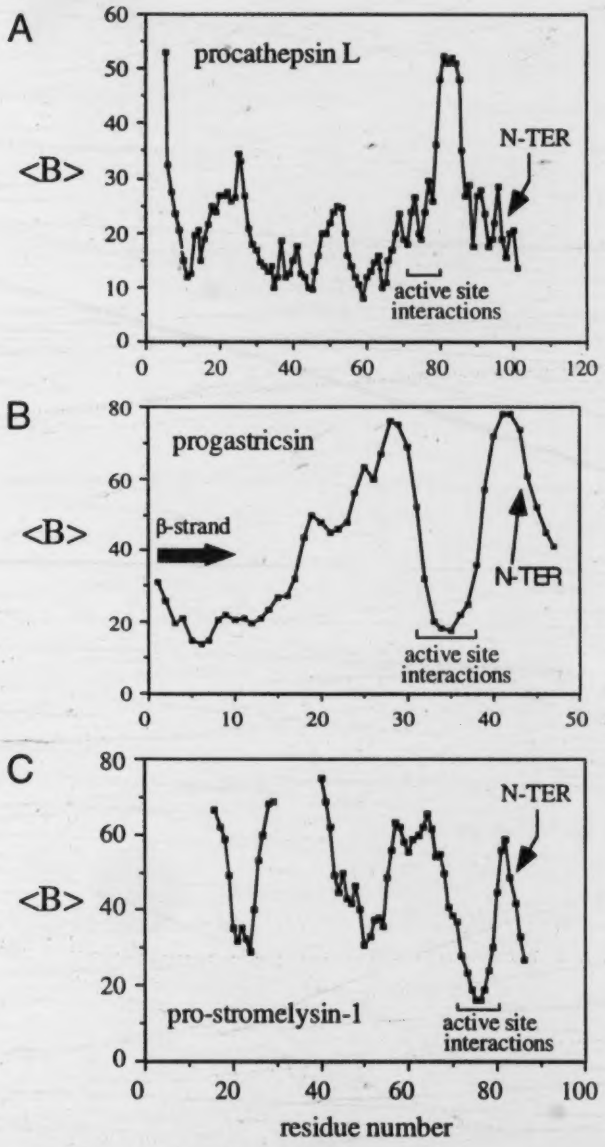


Figure 4B.1: Plots of the mean backbone *B*-factors against the prosegment amino acid sequences of various zymogens. (a) Human procathepsin L, a cysteine proteinase (PDB code 1cjl). The prosegment region at the active site is ordered but is immediately flanked by a region with mobility. (b) Human progastricsin, PDB code 1htr. The prosegment region at the active site has low *B*-factors, but is flanked by regions with high *B*-factors. (c) Human pro-stromelysin-1, a metalloproteinase (PDB code 1slm). Again, the prosegment is stably associated with the active site, but the regions adjacent to the active site have high *B*-factors. The pro-mature junction (N-TER) is indicated in each of the panels.

1994). Examples of serpins include α_1-antitrypsin (the archetypal member), C1-inhibitor (which attenuates the complement cascade), antithrombin, and the 'plasminogen activator inhibitor 1' (PAI-1) that inhibits both the single-chain and two-chain forms of t-PA. The critical balance of proteolytic activity and inhibition is demonstrated by the link between natural variants of serpins and diseases such as emphysema (α_1-antitrypsin), thromboembolic disease (antithrombin) and hereditary angioedema (C1-inhibitor; Stein & Carrell, 1995).

The papain-like cysteine proteinases are strongly but reversibly inhibited by the cystatins, a superfamily of homologous proteins that includes chicken cystatin as the archetypal member (Bode et al., 1988). Small protein-proteinase inhbitors of metalloproteinases are also known (Rees & Lipscomb, 1982; Birkedal-Hansen et al., 1990), such as the 39-residue potato carboxypeptidase inhibitor that may be important for plant defense against fungi (Ryan, 1989). Recently, the first protein-proteinase inhibitor of aspartic proteinases from the intestinal parasitic nematode *Ascaris suum* has been characterized (Martzen et al., 1990). Some of these protein-proteinase inhibitors have been found to mimic substrates, but they subsequently trap the enzyme into a slow and inefficient cleavage reaction. However, the common feature of 'protein-proteinase' inhibitors is that the inactivation mechanisms are distinct from the strategies utilized by prosegments in zymogen structures.

C. Future Directions

The biochemical and structural studies of the past few decades have provided a wealth of insight into the structure and function of enzymes in the pepsin family. However, many new questions have been raised about the aspartic proteinases that will occupy researchers in the years ahead (Blundell, 1998). Although the mammalian, retroviral, and fungal enzymes have been thoroughly characterized by biochemical methods and X-ray crystallography, it is only recently that the plant

aspartic proteinases have been studied in detail. This is somewhat ironic since the pitcher plant 'nepenthes' provided the first hint that the digestive activity of human pepsin was present in other organisms (Blundell et al., 1998). The occasion was a dubious experiment by Charles Darwin in 1875, who noticed that heating of this plant with sulphuric acid "emitted a powerful odour like that of pepsin" (Darwin, 1875). At least 10 aspartic proteinases from plants have now been cloned (Brodelius et al., 1998), and the first crystal structure within this group of enzymes has apparently been determined, although the structure is unpublished. The primary sequences of the plant enzymes contain a large ~100-residue insertion that has no counterpart in other aspartic proteinases, and which may be important for targetting the enzymes to Golgi compartments (Zhu & Conner, 1994).

The pregnancy-associated glycoproteins (PAGs) are pepsin-like proteins that are secreted by placental cells of ruminant species such as cattle and sheep, as well as some non-ruminant species such as pigs (Szafranka et al., 1995). The proteins are specifically targetted to the outer epithelial cells of the placenta, and their function is unknown. Some of these mammals may express 100 or more PAG genes with diverse coding sequences (Xie et al., 1997). The crystal structure of a PAG has not been determined, but a three-dimensional structure has been constructed from the 60% sequence identity to pepsins by comparative molecular modeling (Guruprasad et al., 1996). Although most of the PAGs that have been examined are proteolytically inactive, they may have retained the capacity to bind peptides. In this way, the PAGs may contribute to fetal-maternal communications in these animals (*e.g.*, binding and transporting peptide hormones). Comparisons of the sequences of PAGs with conventional aspartic proteinases, and the absence of PAGs in humans and other mammals, suggest that these genes have evolved recently and rapidly within *Artiodactyla* (Xie et al., 1997). It will be interesting to identify the peptide targets of PAGs and determine their biological role in the developing fetus.

Zymogen activation. Despite the structures of various enzyme species along the activation pathway of the gastric enzymes, it is premature to regard this field as exhausted. In contrast to the gastric enzymes, substantially less is known about the activation mechanisms of other zymogens in the aspartic proteinase superfamily. Furthermore, the energetics of the conversion process are only qualitatively understood. Electrostatic interactions clearly play an essential role as pH sensors to trigger the dramatic conformational changes that are observed during conversion. However, many other residues are also conserved within prosegments (see Fig. 1C.1) which contribute to hydrophobic and hydrogen-bonding interactions, and maintain the secondary structural elements of the prosegments. Very little is known about the relative importance of these residues and their effects upon the conversion process.

One approach to answer these questions would involve the design of an *in vitro* assay system to follow the activation of a recombinant aspartic proteinase, thus allowing mutagenesis of selected residues and subsequent characterization of the effects. This strategy has been employed in a systematic investigation of the positively-charged residues in the prosegment of aspergillopepsin, a pepsin-like enzyme from *Aspergillus niger* (Inoue et al., 1996). It was observed that a Lys residue (K56p) in the 69-residue prosegment was essential for the generation of an active enzyme. The overall results suggested that K56p is important for the proper folding of the zymogen, possibly *via* electrostatic interactions with the active site aspartate residues (analogous to Lys37p of progastricsin). These studies point out a complication in the experimental strategy, namely that problems might exist in differentiating the effects of mutations on the folding of the zymogen *vs.* the subsequent activation pathway.

Aside from these few studies, zymogen activation is a humble companion to the glamorous and lucrative field of structure-based drug design, which has captured attention in the popular media from the success of HIV protease inhibitors like Saquinavir and Ritonavir. As expressed by Thomas Blundell et al. (1998), the attention of the pharmaceutical industry may be regarded as a mixed blessing. One scientifically intriguing area that might overlap with pharmacetical interests regards the activation pathway of the viral aspartic proteinase zymogens. The structural and enzymatic protein components of retroviruses are initially synthesized as a large polyprotein precursor that is cleaved to its functional components by the viral proteinase. However, the proteinase component is initially flanked by other protein domains in the polyprotein from which it must be excised. In fact, the situation is complicated by the fact that HIV protease is functional as a dimer, so that two polyproteins must associate to form the active aspartic proteinase (Co et al., 1994). The emergence of the active HIV protease dimer during maturation of the HIV polyprotein can be considered a special case of zymogen activation. The mechanism is poorly understood and is relevant to human disease because cleavage events by the retroviral protease are an important regulatory mechanism during the life cycle of HIV (Wiegers et al., 1998).

The resources available to modern day laboratories in the areas of molecular biology, biochemistry, structural and computational biology allow a wide range of fundamental questions to be addressed in the field of protein structure and function. It will be interesting to watch the research directions evolve in the aspartic proteinase field.

D. REFERENCES

Bailey, D. & Cooper, J.B. (1994) A structural comparison of 21 inhibitor complexes of the aspartic proteinase from *Endothia parasitica*. *Protein Sci* **3**: 2129-2143.

Birkedal-Hansen, H., Werb, Z., Welgus, H. & Van Wart, H. (1990) In *Matrix Metalloproteinases and Inhibitors*. (Gustav Fischer Verlag, Stuttgart, Germany).

Blundell, T.L., Guruprasad, K., Albert, A., Williams, M., Sibanda, B.L. & Dhanaraj, V. (1998) The aspartic proteinases: a historical overview. *Adv Exp Med Biol* **436**: 1-13.

Bode W. & Huber, R. (1992). Natural protein proteinase inhibitors and their interaction with proteinases. *Eur J Biochem* **204**: 433-451.

Bode, W., Engh, R., Musi, D., Thiele, U., Huber, R., Karshikov, A., Brzin, J., Kos, J. & Turk, V. (1988) The 2.0Å X-ray crystal structure of chicken egg white cystatin and its posssible mode of interaction with cysteine proteinases. *EMBO J* **7**: 2593-2599.

Bode, W., Papamokos, E. & Musil, D. (1987) The high resolution X-ray crystal structure of the complex formed between Subtilisin Carlsberg and eglin-C, an elastase inhibitor from the leech Hirudo medicinalis. *Eur J Biochem* **166**: 673-692.

Brodelius, P.E., Cordeiro, M., Mercke, P., Domingo, A., Clemente, A. & Pais, M.S. (1998) Molecular cloning of aspartic proteinases from flowers of *Cynara cardunculus* subsp. *flavescens* CV. cardoon and *Centaurea calcitrapa*. *Adv Exper Med Biol* **436**: 435-439.

Brunger, A.T. (1992) X-PLOR, a system for X-ray crystallography and NMR. (Yale University Press, New Haven , CT).

Co, E., Koelsch, G., Lin, Y., Ido, E., Hartsuck, J.A. & Tang, J. (1994) Proteolytic processing mechanisms of a miniprecursor of the aspartic protease of human immunodeficiency virus type 1. *Biochemistry* **33**: 1248-1254.

Corey, R.B. & Pauling, L. (1953) Fundamental dimensions of polypeptide chains. *Proc Royal Soc Ser B* **141**: 10-20.

Darwin, C. (1875) *Insectivorous Plants.* (John Murray, London), p. 88.

Dauter, Z., Lamzin, V.S. & Wilson, K.S. (1997) The benefits of atomic resolution. *Curr Op Struct Biol* **7**: 681-688.

Engh, R.A. & Huber, R. (1991) Accurate bond and angle parameters for X-ray protein structure refinement. *Acta Cryst* **A47**: 392-400.

Fraser, M.E., Strynadka, N.C.J., Bartlett, P.A., Hanson, J.E. & James, M.N.G. (1992) Crystallographic analysis of transition-state mimics bound to penicillopepsin: Phosphorus-containing peptide analogues. *Biochemistry* **31**: 5201-5214.

Guruprasad, K., Blundell, T.L., Xie, S.E., Green, J., Szafranska, B., Nagel, R.J., McDowell, K., Baker, C.B. & Roberts, R.M. (1996) *Protein Eng* **9**: 849-856.

Harata, K., Abe, Y. & Muraki, M. (1998) Full-matrix least-squares refinement of lysozyme and analysis of anisotropic thermal motion. *Prot Struct Funct Genet* **30**: 232-243.

Hendrickson, W.A. & Konnert, J.H. (1980) Incorporation of stereochemical information into crystallographic refinement. In *Computing in Crystallography* (Diamond, R., Ramaseshan, S. & Venkatesan, K., eds.), pp. 13.01-13.23, Indian Academy of Sciences, International Union of Crystallography, Bangalore, India.

Inoue, H., Hayashi, T., Huang, X.P., Lu, J.F., Athauda, S.B., Kong, K.H., Yamagata, H., Udaka, S. & Takahashi, K. (1996) Heterologous expression

and site-directed mutagenesis studies on the activation mechanism and the roles of basic residues in the prosegment of aspergillopepsin I. *Eur J Biochem* **237**: 719-725.

James, M.N.G., Sielecki, A.R., Hayakawa, K. & Gelb, M.H. (1992) Crystallographic analysis of transition state mimics bound to penicillopepsin: difluorostatine- and difluorostatone-containing peptides. *Biochemistry* **31**: 3872-3886.

Laskowski, M-Jr. & Kato, I. (1980) Protein inhibitors of proteinases. *Ann Rev Biochem* **49**: 593-626.

Longhi, S., Czjzek, M., Lamzin, V., Nicolas, A. & Cambillau, C. (1997) Atomic resolution (1.0 Å) crystal structure of *Fusarium solani* cutinase: stereochemical analysis. *J Mol Biol* **268**: 779-799.

MacArthur, M.W. & Thornton, J.M. (1996) Deviations from planarity of the peptide bond in peptides and proteins. *J Mol Biol* **264**: 1180-1195.

Martzen, M.R., McMullen, B.A., Smith, N.E., Fujikawa, K. & Peanasky, RJ. (1990) Primary structure of the major pepsin inhibitor from the intestinal parasitic nematode *Ascaris suum*. *Biochemistry* **29**: 7366-7372.

Morris, A.L., MacArthur, M.W., Hutchinson, E.G. & Thornton, J.M. (1992) Stereochemical quality of protein structure coordinates. *Proteins* **12**: 345-364.

Myer, J.S. (1912) In *Life and Letters of William Beaumont* (St. Louis: C.V. Mosby Co.).

Naganathan, P.S. & Venkatesan, K. (1972) Crystal and molecular structure of glycyl-L-alanine hydrochloride. *Acta Cryst* **B28**: 552-556.

Pompers, J.J., Groenewegen, A., de Jong, S., Pepermans, H.A.M. & Hilbers, C.W. (1996) Backbone dynamics of *Fusarium solani pisi* cutinase on four different time scales as determined by NMR. (In Lipases: Structure, Function and Applications: Umea, Sweden), p. 45.

Potempa J., Korzus E. & Travis, J. (1994) The serpin superfamily of proteinase inhibitors: structure, function, and regulation. *J Biol Chem* **269**: 15957-15960.

Ramachandran, G.N. (1968) Need for non-planar peptide units in polypeptide chains. *Biopolymers* **6**: 1494-1496.

Rees, D.C. & Lipscomb, W.N. (1982) Refined crystal structure of potato inhibitor complex of carboxypeptidase A at 2.5Å resolution. *J Mol Biol* **160**: 475-498.

Robertus, J.D., Kraut, J., Alden, R.A. & Birktoft, J.J. (1972) Subtilisin: a stereochemical mechanism involving transition-state stabilization. *Biochemistry* **11**: 4293-4303.

Ryan CA. (1989) Proteinase inhibitor gene families: strategies for transformation to improve plant defenses against herbivores. *Bioessays* **10**: 20-24.

Scapin, G., Gordon, J.I. & Sacchettini, J.C. (1992) Refinement of the structure of recombinant rat intestinal fatty acid-binding apoprotein at 1.2 Å resolution. *J Biol Chem* **267**: 4253-4269.

Scheiner, S. & Kern, C.W. (1977) Theoretical studies of environmental effects on protein conformation. I. Flexibility of the peptide bond. *J Am Chem Soc* **99**: 7042-7050.

Schmitt, E., Mechulam, Y., Fromant, M., Plateau, P., & Blanquet, S. (1997) Crystal structure at 1.2 Å and active site mapping of *Escherichia coli* peptidyl t-RNA hydrolase. *EMBO J* **16**: 4760-4769.

Sheldrick, G.M. (1990) Phase annealing in SHELX-90: direct methods for larger structures. *Acta Crystallogr* **A46**: 467-473.

Sheldrick, G.M. (1997) The SHELX-97 Manual, University of Gottingen, Germany.

Stein P.E. & Carrell R.W. 1995. What do dysfunctional serpins tell us about molecular mobility and disease? *Nature Struct Biol* **2**: 96-113.

Szafranka, B., Xie, S., Green, J. & Roberts, R.M. (1995) Porcine pregnancy-associated glycoproteins: new members of the aspartic proteinase gene family expressed in trophectoderm. *Biol Reprod* **53**: 21-28.

Tame, J.R., Sleigh, S.H., Wilkinson, A.J. & Ladbury, J.E. (1996) The role of water in sequence-independent ligand binding by an oligopeptide transporter protein. *Nature Struct Biol* **3**: 998-1001.

Tsunasawa, S., Masaki, T., Hirose, M., Soejima, M. & Sakiyama, F. (1989) Characteristics of *Achromobacter* protease I, a lysine-specific serine protease. *J Biol Chem* **264**: 3832-3839.

Wang, Z., Luecke, H., Yao, N. & Quiocho, F.A. (1997) A low energy short hydrogen bond in very high resolution structures of protein receptor-phosphate complexes. *Nature Struct Biol* **4**: 519-522.

Wiegers, K., Rutter, G., Kottler, H., Tessmer, H., Hohenberg, H. & Krausslich, H.G. (1998) Sequential steps in immunodeficiency virus particle maturation revealed by alterations of individual Gag polyprotein cleavage sites. *J Virol* **72**: 2846-2854.

Wright, H.T. (1996) The structural puzzle of how serpin serine proteinase inhibitors work. *Bioessays* **18**: 453-464.

Xie, S., Green, J., Bixby, J.B., Szafranska, B., DeMartinini, J.C., Hecht, S. & Roberts, R.M. (1997) The diversity and evolutionary relationships of the pregnancy-associated glycoproteins, an aspartic proteinase sub-family consisting of many trophoblast-expressed genes. *Proc Natl Acad Sci (USA)* **94**: 12809-12816.

Zhu, Y. & Conner, G.E. (1994) Intermolecular association of lysosomal protein precursors during biosynthesis. *J Biol Chem* **269**: 3846-3851.